MADURAI KAMARAJ UNIVERSITY

(University with Potential for Excellence)

DISTANCE EDUCATION

M.B.A

Second Year

IV - Semester

Elective I - Marketing

MARKETING RESEARCH

Recognised by D.E.C.

www.mkudde.org

S417

S417

Madurai Kamaraj University

(University with Potential for Excellence)

Distance Education

www.mkudde.org

ACL - MKU
01122

M.B.A.

SECOND YEAR

IV - Semester

Elective I - Marketing

MARKETING RESEARCH

MADURAI KAMARAJ UNIVERSITY
MADURAI - 625 021.

All copyright privileges are reserved

INTRODUCTION

Market research and marketing research are often confused. *'Market'* research is simply research into a specific market. It is a very narrow concept. *'Marketing'* research is much broader. It not only includes 'market' research, but also areas such as research into new products, or modes of distribution such as via the Internet.

According to American Marketing Association (AMA) "Marketing research is the function that links the consumer, customer, and public to the marketer through information - information used to identify and define marketing opportunities and problems; generate, refine, and evaluate marketing actions; monitor marketing performance; and improve understanding of marketing as a process. Marketing research specifies the information required to address these issues, designs the methods for collecting information, manages and implements the data collection process, analyzes, and communicates the findings and their implications."

The purpose of this elective paper is to make the students aware of the nature scope and importance of marketing research with special reference to marketing environment. The elaborate coverage of techniques in marketing research will help students appreciate the subject and encourage them to do projects using these techniques.

Printed at Reva - C / 2000 - 2013

SYLLABUS

Introduction to marketing research – Marketing research as a tool of management – Relevance of Marketing research in the Indian Marketing environment.

Basic concepts – Scientific method – Types of research – basic methods of collecting data- Secondary data.

The marketing research process – Planning the research project. The data collection forms – Attitude measurements – Introduction to sampling – Applications of sampling methods to marketing Problems – Data collection and the field force – Tabulatin of collected data – Analysis – Techniques – Research report presentations.

Application of Marketing research

a) Identifying Market Segments

b) Product Research

c) Advertising Research

d) Market and Sales Analysis Research

CONTENTS

Unit I
INTRODUCTION TO MARKETING RESEARCH 1-24

Unit II
MARKETING RESEARCH PROCESS 25 - 41

Unit III
RESEARCH DESIGN 42 – 66

Unit IV
BASIC METHODS OF COLLECTING DATA 67 – 96

Unit V
SAMPLING DESIGN 97 – 120

Unit VI
SCALES AND ATTITUDE MEASUREMENT 121 – 155

Unit VII
DATA PROCESSING 156 – 186

Unit VIII
UNIVARIATE HYPOTHESIS TESTING AND BIVARIATE ANALYSIS 187 – 216

Unit IX
MULTIVARIATE DATA ANALYSIS 217 – 242

Unit X
APPLICATIONS OF MARKETING RESEARCH 243 – 280

UNIT I

INTRODUCTION TO MARKETING RESEARCH

STRUCTURE

1.1 Introduction to Marketing Research
1.2 Definitions of Marketing Research
1.3 Objectives of Marketing Research
1.4 Broad types of Marketing Research
1.5 Application based Marketing Research types
1.6 Importance of Marketing Research
1.7 Benefits of Marketing Research
1.8 Scope of Marketing Research
1.9 Difference between Market Research and Marketing Research
1.10 Factors affecting use of Marketing Research
1.11 Limitations of Marketing Research
1.12 Characteristics of good Marketing Research
1.13 Summary
1.14 Answers to 'Check Your Progress'
1.15 Exercises and Questions
1.16 Further Readings

UNIT OBJECTIVES

After reading this unit, you will be able to

- Understand the applications and importance of marketing research
- Describe and explain the impact marketing research has on marketing decisions.
- Understand the various classifications of Marketing Research
- Differentiate marketing research from market research

Notes

> ➢ Understand the scope and limitations of the marketing research.

1.1 INTRODUCTION TO MARKETING RESEARCH

The field of marketing research as a statistical science was pioneered by Arthur Nielsen with the founding of the AC Nielsen Company in 1923. Marketing research is a systematic and objective study of problems pertaining to the marketing of goods and services. Research is the only tool an organization has to keep in contact with its external operating environment. Marketing is the process of planning and executing the pricing, promotion, and distribution of products, services, and ideas in order to create exchanges that satisfy both the firm and its customers. The fundamental purpose of marketing is to allow a firm to plan and execute the pricing, promotion, and distribution of products, services, and ideas in order to create exchanges that satisfy both the firm and its customers. Marketing research is the mechanism for generating that information. Marketers require following market information:

- Who are the customers?
- Where are they located and how can they be contacted?
- What quantity and quality do they want?
- What is the best time to sell?
- What is the long-term or historical price data over a number of years?
- What is the expected production in the country?
- Is there more demand for one product or another?

Check your Progress

1. What are the two types of research?

1.1.1 Marketing Research

Marketing research is undertaken to assist the marketing function. Marketing research stimulates the flow of marketing data from the consumer and his environment to marketing

information system of the enterprise. Market research involves the process of

- Systematic collection
- Compilation
- Analysis
- Interpretation of relevant data for marketing decisions

This information goes to the executive in the form of data. On the basis of this data the executive develop plans and programmers. Advertising research, packaging research, performance evaluation research, sales analysis, distribution channel, etc., may also be considered in management research. Research tools are applied effectively for studies involving:

1. Demand forecasting
2. Consumer buying behaviour
3. Measuring advertising effectiveness
4. Media selection for advertising
5. Test marketing
6. Product positioning
7. Product potential

1.2 DEFINITIONS OF MARKETING RESEARCH

Authors have defined Marketing Research in many ways:

Kotler (1999) defines marketing research as 'systematic problem analysis, model-building and fact-finding for the purpose of improved decision-making and control in the marketing of goods and services'.

Kotler and keller (2006) define marketing research as systematic design, collection, analysis, and reporting of data and findings relevant to a specific marketing situation facing the company.

The American Marketing Association (AMA, 1961) defines it as 'the systematic gathering, recording and

Notes

Check your Progress

2. What is a basic research?

Notes

analyzing of data relating to the marketing of goods and services'.

Green and Tull have defined marketing research as the systematic and objective search for and analysis of information relevant to the identification and solution of any problem in the field of marketing.

The American Marketing Association formally defines marketing research as follows: Marketing research is the function that links an organization to its market through the gathering of information. This information allows for the identification and definition of market-driven opportunities and problems.

1.3 OBJECTIVES OF MARKETING RESEARCH

The objectives of marketing research generally are:

- To generate, refine, evaluate marketing plan.
- To monitor marketing performance and improve impact of marketing programme.
- To identify market potential
- To assess the stress on innovation for market growth and profitability.

1.4 BROAD TYPES OF MARKETING RESEARCH

Broadly marketing research is classified into

1. Exploratory Research
2. Descriptive Research
3. Causal Research

Let us see each of them one by one

Check your Progress

3. What is an applied research?

1.4.1 Exploratory Research

Exploratory research focuses on collecting either secondary or primary data and using an unstructured format or informal procedures to interpret them. Among the three types of research designs, exploratory research incorporates the fewest

Notes

characteristics or principles of the scientific method. It is often used simply to classify the problems or opportunities and it is not intended to provide conclusive information from which a particular course of action can be determined. Some examples of exploratory research techniques are focus group interviews, experience surveys, and pilot studies. Exploratory research may also use some forms of secondary data (e.g., online databases). Exploratory research can be somewhat intuitive and is used by many decision makers who monitor market performance measures pertinent to their company or industry.

Many times a decision maker is grappling with broad and poorly defined problems. If you attempt to secure better definitions by analytic thinking, it may be the wrong approach and may even be counter productive counter productive in the sense that this approach may lead to a definitive answer to the wrong question. Exploratory research uses a less formal approach. It pursues several, possibilities simultaneously and in a sense it is not quite sure of its objective. Exploratory research is designed to provide a background, to familiarize and, as the word implies, just "explore", the general subject. A part of exploratory research is the investigation of relationships among variables without knowing why they are studied. It borders on an, idle curiosity approach, differing from it only in that the investigator thinks there may be a payoff in the application some where in the forest of questions. Three typical approaches in exploratory research are:

a. The literature survey,

b. The experience survey, and

c. The analysis of "insight-stimulating" examples

The literature search is fast, economical way to develop a better understanding of a problem area in which you are

Check your Progress

4. What is meant by qualitative research?

Notes

investigating and have limited experience and knowledge. It also familiarizes you with past research results, data sources, and the type of data available. The experience survey concentrates on persons who are particularly knowledgeable in the particular area. In this representative samples are not desired. A covering of widely divergent views is better. Researchers are not looking for conclusions; they are looking for ideas. The analysis of specific examples is a sort of case study approach, but again, researchers are looking for some fresh possible divergent views.

1.4.2 Descriptive Research

Descriptive research as the name suggests is designed to describe something- for example, the characteristics of users of a given product; the degree to which product use varies with income, age, sex or other characteristics; or the number who saw a specific television commercial. To be of maximum benefit, a descriptive study must only collect data for a definite purpose. Your objective and understanding should be clear and specific. Descriptive studies vary in the degree to which a specific hypothesis is the guide. It allows both implicit and explicit hypotheses to be tested depending on the research problem. For Example: A cereal company may find its sales declining. On the basis of market feedback the company may hypothesize that teenage children do not eat its cereal for breakfast. A descriptive study can then be designed to test this hypothesis.

Descriptive research uses a set of scientific methods and procedures to collect raw data and create data structures that describe the existing characteristics (e.g., attitudes, intentions, preferences, purchase behaviors, evaluations of current marketing mix strategies) of a defined target population or market structure. Descriptive research designs are appropriate

when the research objectives include determination of the degree to which marketing (or decision) variables are related to actual market phenomena. Here, the researcher looks for answers to the how, who, what, when, and where questions concerning different components of a market structure. Descriptive studies generally allow decision makers to draw inferences about their customers, competitors, target markets, environmental factors, or other phenomena of concern. For example, there is a growing trend among today's major chain restaurants to conduct annual studies designed to identify and describe consumers' attitudes, feelings, and patronage behavior toward their own restaurants as well as toward those of their main competitors. These studies, referred to as either image assessment surveys or customer service satisfaction surveys, partially describe how consumers rate different restaurants' customer service, convenience of location, food quality, overall quality, and so on. However, descriptive designs are not capable of addressing any of the why questions associated with a given research problem. Still, the data and information generated through descriptive designs can provide decision makers with evidence that can lead to a course of action.

Unlike exploratory studies, descriptive studies are characterized by a formal design and an accurate description of the problem. This helps in identifying the information required and ensures that it covers all the areas required. It is imperative that the design of descriptive studies be such that it specifies the source of information and the data to be collected from those sources. This is done mainly to ensure the accuracy and the appropriateness of the information collected. It is equally necessary to prevent the collection of any unnecessary

data associated with such research. There are basically two types of descriptive studies.

1. Case method
2. Statistical method

1.4.2.1 Case Method

This method is not often used in descriptive research. It is more widely used in exploratory research. The procedure is the same as in exploratory research. The only difference between the two lies in the fact that while exploratory research offers flexibility, descriptive research is more structured, clearly defining the research problem and the points to be investigated.

1.4.2.2 Statistical Method

This is the most widely used method in Marketing Research. It makes use of techniques that vary from simple means and percentages to very sophisticated techniques. In this method, a limited number of factors from a large number of cases are studied in-depth. Statistical tools are used by most marketing research professionals to understand the dynamics of the market. Data is usually collected through observation or through interviewing. Surveys to find the consumption patterns of children ad generate profiles of the people using the internet to make purchases are examples of the statistical method. Let us see how Forrester research, the leading global marketing research agency, has studied online purchases. They have discovered that the average online buyer reports purchases in two product categories: books and software. The following are some of the observations made by Forrester on line purchases.

· Media and technology lead net shopping. More than 50% of online shoppers buy software and more than 40% buy books.

Buyers spend an a ge of more than $100 over three- month period.

· More than 250,000 North Americans bought cars and computers online.

· Three percent of online consumers report online securities trades.

Advantages of Statistical Method

Statistical study involves a large number of interviews or observation. Statistical techniques are used specifically for mass data. Two different researchers conducting statistical research will arrive at the same results, while two people using the case study method may not arrive at the same results. This is mainly because of the subjectivity of the case study method. Statistical study helps the researcher to make more accurate generalizations. If the sampling is properly done, the generalization will be universally true.

Disadvantages

· Fails to prove cause-and-effect relationship

· Direction of causal effect is not clear in statistical studies.

1.4.3 Causal Research

Causal research is designed to collect raw data and create data structures and information that will allow the decision maker or researcher to model cause-and-effect relationships between two or more market (or decision) variables. Causal research is most appropriate when the research objectives include the need to understand the reasons why certain market phenomena happen as they do. That is to say, the decision maker may have a strong desire to understand which (independent) market or decision variables are the cause of the dependent phenomenon defined in the decision or research problem. Causal research can be used to understand the functional relationships

Notes

Check your Progress

5. What is meant by quantitative research?

Notes

between the causal factors and the effect predicted on the market performance variable under investigation. In addition, understanding the cause–effect relationships among market performance factors allows the decision maker to make if–then statements about the variables. For example, the owner of a small casual men's clothing store in Madurai might be able to say, "If I expand the assortment of brand-name shirts, increase my advertising budget by 15 percent, have an introductory 30 percent–off sale on the new shirts, and keep the rest of my marketing mix strategies unchanged, then our overall sales volume will increase by 40 percent." While causal research designs offer opportunity for identifying, determining, and explaining causality among critical market factors, they tend to be complex, expensive, and time-consuming. Among the different data collection techniques available, experimental designs hold the greatest potential for establishing cause–effect relationships because they allow researchers to investigate changes in one variable while manipulating one or two other variables under controlled conditions. Causal research is also a technique used to perform conclusive research. It attempts to specify the nature of the functional relationship between two or more variables in the problem model. Managers analyze the impact of advertising etc. through causal research. For instance, once they give an advertisement they study how the advertising has caused the sales to change. Usually conclusions are drawn from three types of evidence.

- Concomitant variations
- Sequence of concurrence
- Absence of other potential casual factors.

Concomitant Variations- Advertising expenditures for an organization vary across geographic regions. An organization

may notice the variations of sales across the region. It may also notice that sales revenue is high in those regions where the advertising expenditures are high and is low in those regions where advertisement expenditures are low. So it may infer that sales revenue is directly proportional to advertisement expenditure. But this has only been inferred, not proved.

Sequence of Occurrence- This is another type of evidence that can be used to make inferences about causation. Some events that happen first cause the next event to occur. In such cases, researchers can infer that the first event has caused the second event.

Absence of Other Potential Factors- If the causative factors are identified the researcher will be able to eliminate all the factors except the one that he believes is the real causative factor. Then he can establish the same as the real causative factor. Quite often it is difficult to eliminate all the possible causative factors.

1.5 CLASSIFICATION OF MARKETING RESEARCH BASED ON APPLICATION

Marketing research is classified based on the application as follows:

1. Product Research: Assessment of suitability of goods with respect to design and price.

2. Market Characteristics Research (Qualitative): Who uses the product? Relationship between buyer and user, buying motive, how a product is used, analysis of consumption rates, units in which product is purchased, customs and habits affecting the use of a product, consumer attitudes, shopping habits of consumers, brand loyalty, research of special consumer groups, survey of local markets, basic economic analysis of the consumer market, etc.

Notes

3. *Size of Market (Quantitative)*: Market potential, total sales quota, territorial sales quota, quota for individuals, concentration of sales and advertising efforts; appraisal of efficiency, etc.

4. *Competitive position and Trends Research:*

5. *Sales Research: Analysis of sales records:*

6. *Distribution Research:* Channels of distribution, distribution costs.

7. *Advertising and Promotion Research:* Testing and evaluating, advertising and promotion

8. *New product launching and Product Positioning Research:*

9. *Ad Tracking:* periodic or continuous in-market research to monitor a brand's performance using measures such as brand awareness, brand preference, and product usage.

10. *Advertising Research:* used to predict copy testing or track the efficacy of advertisements for any medium, measured by the ad's ability to get attention, communicate the message, build the brand's image, and motivate the consumer to purchase the product or service.

11. *Brand equity research:* how favorably do consumers view the brand?

12. *Brand association research* - what do consumers associate with the brand?

13. *Brand attributes research* - what are the key traits that describe the brand promise?

14. *Brand/ name testing* - what do consumers feel about the names of the products?

15. *Commercial eye tracking research* - examine advertisements, package designs, websites, etc by analyzing visual behavior of the consumer.

Notes

16. *Concept testing* - to test the acceptance of a concept by target consumers

17. *Cool hunting* - to make observations and predictions in changes of new or existing cultural trends in areas such as fashion, music, films, television, youth culture and lifestyle

18. *Buyer decision processes research* - to determine what motivates people to buy and what decision-making process they use

19. *Copy testing* – predicts in-market performance of an ad before it airs by analyzing audience levels of attention, brand linkage, motivation, entertainment, and communication, as well as breaking down the ad's flow of attention and flow of emotion.

20. *Customer satisfaction research* - quantitative or qualitative studies that yields an understanding of a customer's of satisfaction with a transaction

21 *Demand estimation* - to determine the approximate level of demand for the product

22. *Distribution channel audits* - to assess distributors' and retailers' attitudes toward a product, brand, or company

23. *Internet strategic intelligence* - searching for customer opinions in the Internet: chats, forums, web pages, blogs... where people express freely about their experiences with products, becoming strong "opinion formers"

24. *Marketing effectiveness and analytics* - Building models and measuring results to determine the effectiveness of individual marketing activities.

25. *Mystery Consumer or Mystery shopping* - An employee or representative of the market research firm anonymously contacts a salesperson and indicates he or she is shopping for a product. The shopper then records the entire experience. This method is

often used for quality control or for researching competitors' products.

26. *Positioning research* - how does the target market see the brand relative to competitors? - what does the brand stand for?

27. *Price elasticity testing* - to determine how sensitive customers are to price changes

28. *Sales forecasting* - to determine the expected level of sales given the level of demand. With respect to other factors like Advertising expenditure, sales promotion etc.

29. *Segmentation research* - to determine the demographic, psychographic, and behavioural characteristics of potential buyers

30. *Online panel* - a group of individual who accepted to respond to marketing research online

31. *Store audit* - to measure the sales of a product or product line at a statistically selected store sample in order to determine market share, or to determine whether a retail store provides adequate service

32. *Test marketing* - a small-scale product launch used to determine the likely acceptance of the product when it is introduced into a wider market

33. *Viral Marketing Research* - refers to marketing research designed to estimate the probability that specific communications will be transmitted throughout an individuals Social Network. Estimates of Social Networking Potential (SNP) are combined with estimates of selling effectiveness to estimate ROI on specific combinations of messages and media.

1.6 IMPORTANCE OF MARKETING RESEARCH

Marketing Research helps the marketer in following decision areas:

1.6.1 Target Market

To understand taste, preferences and choices of consumers, to understand market size, to measure market potential, to interpret consumer behavior, to study influence of life style on target market behavior.

1.6.2 Products / Services

To identify customer satisfaction, to identify customer service levels, to study and to augment product features, attributes as well as to identify service gaps

1.6.3 Price

To study price affordability of target market, to study competitive pricing structure, etc.

1.6.4 Distribution

To identify prevailing channels of distribution, emerging channels, channels of distribution by competitors and modification to be done in channels structures as per market requirement.

1.6.5 Promotion

To design promotion mix, to identify promotion mix of competitors, to study emerging promo tools (road shows, pops and kiosks)

1.7 BENEFITS OF MARKETING RESEARCH

The benefits of marketing research are it helps

- to identify strength and weaknesses of the marketer as well as the competitors,
- to research market shares,
- to determining whether economics of scale and economics of scope (e.g. retailing, insurance and telecom

can be clubbed by organized retailer like Shopper's stop, ICICI, etc.),

➤ to understand needs wants and demands of target customers in India and abroad,

➤ to formulate sales and distributions strategies, and

➤ to avoid business recession by timely launching brand extension and or product variants (detail study of PLC or BCG matrix).

1.8 SCOPE OF MARKETING RESEARCH

Marketing research has a vast scope it can be used in:

1.8.1 Product Management: One of the major scope of marketing research is to manage the current products and new products. In product management Marketing Research is helpful in

1. Competitive Intelligence – To understand the competitive product stretegy.
2. Pre-launch strategy for new products Test Marketing – To monitor the performance of the brand by launching in a select area and then taking it across the country. In other words it is a small-scale product launch used to determine the likely acceptance of the product when it is introduced into a wider market.
3. Concept testing - to test the acceptance of a concept by target consumers.

1.8.2 Sales analysis: Marketing research is used to study the sales trend and make suitable strategies when required. It is used to

a. Assess market potential

b. Estimation of demand for a product

c. Market share estimation

d. Study seasonal variation for a product

Notes

e. Market segmentation studies

f. Estimate size of the market

g. Need analysis to find out where the product fits in

1.8.3 Corporate Research: Marketing Research is used to analyse the corporate effectiveness. Some examples are:

1. Assessing the image of the company

2. Knowledge of the company activities

1.8.4 Advertising Research: Advertising is an arena in which Marketing Research is extensively used. Some scopes are:

1. Readership feedbacks – Mainly carried out for newspapers and magazines

2. Advertising Recall – To assess the recall of telivision or other advertising and thereby assess its effectiveness.

1.9 DIFFERENCE BETWEEN MARKET RESEARCH AND MARKETING RESEARCH

Generally Market Research and Marketing Research are confused to be the same. But there is a clear distinction between the both.

Market Research: Market Research involves researching a specific industry or market. Ex: Researching the automobile industry to discover the number of competitors and their market share.

Marketing Research: Marketing Research analyses a given marketing opportunity or problem, defines the research and data collection methods required to deal with the problem or take advantage of the opportunity, through to the implementation of the project. It is a more systematic method which aims to discover the root cause for a specific problem within an organisation and put forward solutions to that problem. Ex: Research carried out to analyze and find solution for increasing turnover in an organisation.

Notes

1.10 FACTORS AFFECTING USE OF MARKETING RESEARCH

In spite of the rapid growth of marketing research, many companies still fail to use it sufficiently or correctly. Several factors can stand in the way of its greater utilization.

A narrow conception of marketing research. Many managers see marketing research as only a fact-finding operation. The marketing researcher is supposed to design a questionnaire, choose a sample conduct interviews, and report results, often without being given a careful definition of the problem or of the decision alternatives facing management. As a result, some fact finding fails to be useful. This reinforces management's idea of the limited usefulness of some marketing research.

Uneven caliber of marketing researchers. Some managers view marketing research as little better than a clerical activity and reward it as such. Poorly qualified marketing researchers are hired, and their weak training and deficient creativity lead to unimpressive results. The disappointing results reinforce management's prejudice against expecting too much from marketing research. Management continues to pay low salaries, perpetuating the basic difficulty.

Late and occasional erroneous findings by marketing research. Managers want quick results that are accurate and conclusive. But good marketing research takes time and money. If they can't perceive the difference between quality and shoddy research, managers become disappointed, and they lower their opinion of the value of marketing research. This is especially a problem in conducting marketing research in foreign countries.

Intellectual differences. Intellectual divergences between the mental styles of line managers and marketing researchers often get in the way of productive relationships. The marketing

Notes

researcher's report may seem abstract, complicated, and tentative, while the line manager wants concreteness, simplicity, and certainty. Yet in the more progressive companies, marketing researchers are increasingly being included as members of the product management team, and their influence on marketing strategy in growing.

1.11 LIMITATIONS OF MARKETING RESEARCH

Corporate India's turnover is few billion $ whereas sales turnover of all marketing research companies (organize and unorganized) does not exceed more than Rs. 500 crores. This indicates that marketing research is not very popular with corporate India. The big shots in consumer non-durables i.e HLL and ITC hardly spend around Rs. 25 crores and 20 crores respectively annually on marketing research, which is not even 1 per cent of their sales turnover. This fact confirms unpopularity of marketing research with corporate world. The reasons for unpopularity could be as follows:

a. **Narrow conceptions:** MR is perceived as data collection activity only i.e. a clerical job. This is because the marketer never comes in forward of respondents and explains to him the objectives of research or purpose of research. E.g. Times of India and Indian Express conduct research for estimating readership and viewer ship every six months. The boys recruited are under-graduates who do not know the purpose of the study. They judge jott down the answers on structured questionnaires and say that they are doing marketing research. Hence the respondents form perception that marketing research is clerical job.

b. **Improper orientation of the investigators** – data collection activity is normally enthrusted to first year management

students without imparting any training to them. The respondents are just given quota i.e. they have to complete say 100 surveys in one week's time and submit 100 questionnaires. Normally, the interviewer is not able to complete this work in one week's time and to fulfill the quota on his own he fills up the questionnaire. This hampers the accuracy of the survey.

c. Less number of marketing research firms: In corporate India total number of companies could be more than 10,000 whereas companies engage in conducting marketing research, in organised sector are around 10 to 15 and unorganized around 32 to 50; of which the major are leading marketing research companies and their sales turnover is as follows:

MARKETING RESEARCH COMPANY	SERVICES OFFERED	SALES TURNOVER FOR THE YEAR 2005 (RS. IN CRORES)
ORG-MARG	People meter, retail audit	100
MRAS	Test marketing	25
MBA	Opinion polls	10
RCG	Perceptual maps	12
STANDARD RESEARCH	Customer satisfaction surveys	10
IMRB	Advertising testing research	75
	Total	232

d. Late results - well design and plan survey which is to be completed by conducting personal interviews might take 4 to 6 months time. In marketer's opinion, the survey should not take more than one month's time since he perceives it as clerical job.

As such the report submitted by marketer may not be attended by the sponsors.

e. Conditional findings – MR companies normally want to play safe i.e. due to volatile Indian markets, they never recommend any marketing strategy. In place they normally recommend conditional strategies i.e. if this happen, this marketing strategy will work. Such conditional marketing plans are not acceptable to the marketer because marketer can hardly control the conditions.

f. Cost affair – Field research is always very costly because the expenses like traveling, conveyance, lodging, meals, communication, etc. to be incurred e.g. a survey done for nation wide market for consumer non-durable like toothpaste might require few crore rupees. However, marketers opinion is it should not take more than few thousand since he perceives MR as clerical job.

g. Biasness – Research may have biased due to

Improper research techniques

Inadequate skill of investigator

Researcher inclined towards pre-determined results.

1.12 CHARACTERISTICS OF GOOD MARKETING RESEARCH

Following are the characteristics of good marketing research

a. Scientific method: Effective marketing research uses the principles of the scientific method: careful observation, formulation of hypotheses, prediction, and testing.

b. Research creativity: At its best, marketing research develops innovative ways to solve a problem.

c. Multiple methods: Competent marketing researchers shy away from over-reliance on any one method, preferring to adapt the method to the problem rather than the other way

around. They also recognize the desirability of gathering information from multiple sources to give greater confidence.

d. Interdependence of models and data: Competent marketing researchers recognize that the facts derive their meaning from models of the problem. These models guide the type of information sought and therefore should be made as explicit as possible.

e. Value and cost of information: Competent marketing researchers show concern for estimating the value of information against its cost. Value/cost evaluation helps the marketing research department determine which research projects to conduct, which research designs to use, and whether to gather more information after the initial results are in. Research costs are typically easy to quantify, while the value is harder to anticipate. The value depends on the reliability and validity of the research findings and management's willingness to accept and act on its findings. In general, the most valuable information tends to cost the most because it requires more intensive methods, but of course it is easy to spend a great deal of money on poorly conceived research.

f. Healthy skepticism: Competent marketing researchers will show a healthy skepticism toward assumptions made by managers about how the market works.

g. Ethical marketing: Most marketing research benefits both the sponsoring company and its consumers. Through marketing research, companies learn more about consumers' needs, and are able to supply more satisfying products and services. However, the misuse of marketing research can also harm or annoy consumers. There are professional ethical standards guiding the proper conduct of research.

Notes

1.13 SUMMARY

Thus we are able to understand the concept of marketing research, its scope and limitations.

1.14 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The two types of research are basic and applied research.
2. Basic research is the pure research which is more educative.
3. Applied research is carried out to seek alternate solution for a problem. Applied research can further be classified as problem-solving research and problem-oriented research depending upon the research problem at hand.
4. Qualitative marketing research – This is generally used for exploratory purposes. The data collected is qualitative and focuses on people's opinions and attitudes towards a product or service. The respondents are generally few in number and the findings cannot be generalised to the whole population. No statistical methods are generally applied. Ex: Focus groups, In-depth interviews, and Projective techniques
5. Quantitative marketing research – This is generally used to draw conclusions for a specific problem. It tests a specific hypothesis and uses random sampling techniques so as to infer from the sample to the population. It involves a large number of respondents and analysis is carried out using statistical techniques. Ex: Surveys and Questionnaires

1.15 EXERCISES AND QUESTIONS

1. What is Marketing Research?
2. Define Marketing Research as stated by various authors.
5. List down the scope of Marketing Research.
6. What is the difference between problem-solving research and problem- oriented research?
7. Give an example of situations where you will use exploratory research.

Notes

1.16 FURTHER READING

1. Boyd Harper W., Ralph Westfall and Stanley F Stasch. Marketing Research Text and Cases, Irwin, Inc., 2007

2. Green, Paul E. and Donald S. Tull, Research For Marketing Decisions, Englewood clifs N.J, Prentice Hall, Inc. 2006

3. David.J.Lick and Donald.S.Rubin, Marketing Research, 7th edition, PHI, 2007.

4. G.C.Beri, Marketing Research, TMH, 2008.

5. Nargundkar R, Marketing Research Text and Cases, TMH, 2002

6. Dr.D.D.Sharma, Marketing research, Principles, applications, and cases, Sultan Chand and sons, 2007.

UNIT II

THE MARKETING RESEARCH PROCESS

STRUCTURE

2.1 Introduction

2.2 Marketing Research process

2.3 Errors in Marketing Research process

2.4 Summary

2.5 Answers to 'Check Your Progress'

2.6 Exercises and Questions

2.7 Further Readings

UNIT OBJECTIVES

After learning this unit you would be able

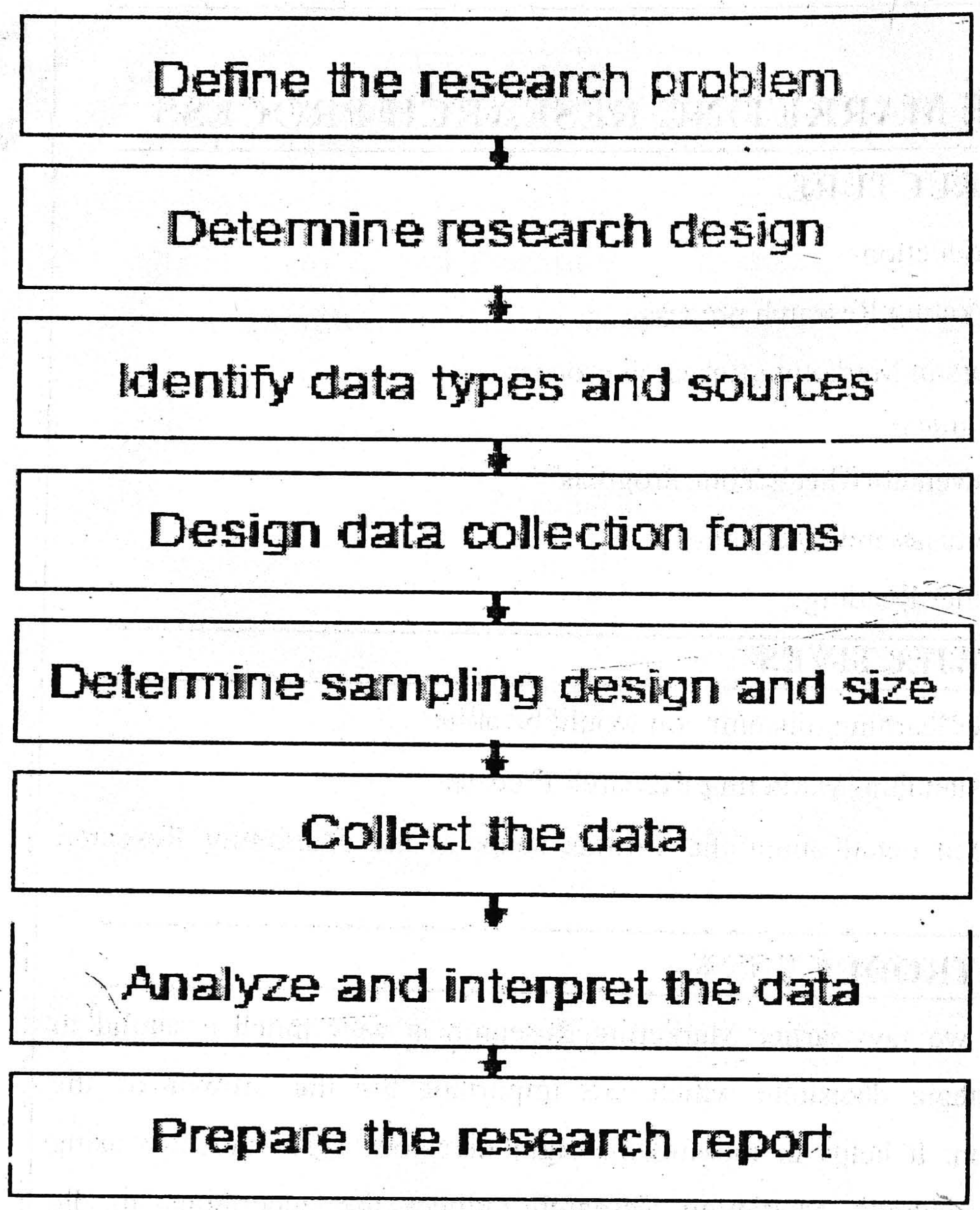
- To understand the Marketing Research Process.
- To learn in detail about the various steps in the Marketing Research Process

2.2 INTRODUCTION

As we saw earlier Marketing Research is very much essential to make strategic decisions which are important for the growth of the organisation. It helps in making the right decisions systematically using statistical methods. Marketing Research reduces the uncertainty in the decision-making process and increase the probability and magnitude of success if conducted in a systematic, analytical, and objective manner. Marketing research by itself does not arrive at marketing decisions, nor does it guarantee that the organization will be successful in marketing its products. It is only a tool which helps in the decision making process.

2.2 THE MARKETING RESEARCH PROCESS

The Marketing Research Process involves a number of inter-related activities which have bearing on each other. Once the need for Marketing Research has been established; broadly it involves the steps as depicted in Figure below:



Let us now know in detail about the various steps involved in the Marketing Research Process

1. Define the Research Problem

The first step in Marketing is to define the research problem. A problem well defined is half-solved. If a problem is poorly defined, a good research design cannot be developed. The decision problem faced by the organisation must be translated into a market research problem in the form of questions. These questions must define the information that is required to make the decision and how this information can be obtained. This way, the decision problem gets translated into a research problem. For example, a decision

problem may be whether to launch a new product. The corresponding research problem might be to assess whether the market would accept the new product. In order to define the problem more precisely, an exploratory research can be carried out. Survey of secondary data, pilot studies or experience surveys are some of the popular methods.

The step defining the research problem exists of 2 main steps: (1) formulating the problem and (2) establishing research objectives.

Defining the problem is the single most important step in the market research process. A clear statement of the problem is a key to a good research. A firm may spend hundreds or thousands of dollars doing market research, but if it has not correctly identified the problem, those dollars are wasted. In our case it is obvious that the problem here is setting up a business. But even if this is clear, you still need to know what exactly you need to know to make the new business a success and what specific related to the product is difficult to find out. Problems that may be encountered are: it is unknown what potential markets there are, what customer groups are interested in your products, who the competitors are? After formulating your problem, you need to formulate your research questions. What questions need to be answered and which possible sub-questions do you have.

With the problem or opportunity defined, the next step is to set objectives for your market research operations. Research objectives, related to and determined by the problem formulation, are set so that when achieved they provide the necessary information to solve the problem. A good way of setting research objectives is to ask, "What information is needed in order to solve the problem?" Your objective might be to explore the nature of a problem so you may further define it, or perhaps it is to determine how many people will buy your product packaged in a certain way and offered at a certain price. Your objective might even be to test

possible cause and effect relationships. For example, if you lower your price, how much will it increase your sales volume? And what impact will it have on your profit? Clear objectives can lead to clear results. An example of this is a situation at Camaro/Firebird. Auto manufacturers are sometimes criticized for creating expensive vehicles with unwanted features and technologies that do not meet the needs of the target market. To avoid this trap engineering team of this company turned to market research to evaluate how changes in performance and fuel economy would affect sales volume and customer satisfaction. It turned out that customers were willing to pay more for greater performance if the car also offered simultaneous increases in fuel economy. The problem description, the research question, sub questions and the research objectives are part of an overall document problem description. After describing and formulating the problem and the objectives, the next step is to prepare a detailed and realistic time frame to complete all steps of the market research process. If your business operates in cycles, establish target dates that will allow the best accessibility to your market. For example, a holiday greeting card business may want to conduct research before or around the holiday season buying period, when their customers are most likely to be thinking about their purchases.

2. Determine research design

The research design specifies the method and procedure for conducting a particular study. The step selecting and establishing research design consists of 3 main steps: (1) select the research design, (2) identify information types and sources and (3) determine and design research instrument.

As stated earlier, every research project and every business is different. Still, there are enough commonalities among research projects to categorize them by research methods and procedures used to collect and analyze data. As studied already, marketing research and hence the research designs can be classified into one of three categories

- Exploratory research
- Descriptive research
- Causal research

This classification is based on the objective of the research. In some cases the research will fall into one of these categories, but in other cases different phases of the same research project will fall into different categories. Problems are formulated clearly in exploratory research. It aims at clarifying concepts, gathering explanations, gaining insight, eliminating impractical ideas, and forming hypotheses. Exploratory research can be performed using a literature search, surveying certain people about their experiences, focus groups, and case studies. During the survey, exploratory research studies would not try to acquire a representative sample, but rather, seek to interview those who are knowledgeable and who might be able to provide insight concerning the relationship among variables. Case studies can include contrasting situations or benchmarking against an organization known for its excellence. Exploratory research may develop hypotheses, but it does not seek to test them. Exploratory research is characterized by its flexibility. A descriptive study is undertaken when the researcher wants to know the characteristics of certain groups such as age, sex, educational level, income, occupation, etc. Descriptive research is more rigid than exploratory research and seeks to describe users of a product, determine the proportion of the population that uses a product, or predict future demand for a product. Descriptive research should define questions, people surveyed, and the method of analysis prior to beginning data collection. In other words, the who, what, where, when, why, and how aspects of the research should be defined. Such preparation allows one the opportunity to make any required changes before the costly process of data collection has begun. There are two basic types of descriptive research: longitudinal studies and cross sectional studies. Longitudinal

NOTES

studies are time series analyses that make repeated measurements of the same individuals, thus allowing one to monitor behavior such as brand switching. However, longitudinal studies are not necessarily representative since many people may refuse to participate because of the commitment required. Cross-sectional studies sample the population to make measurements at a specific point in time. A special type of cross-sectional analysis is a cohort analysis, which tracks an aggregate of individuals who experience the same event within the same time interval over time. Cohort analyses are useful for long-term forecasting of product demand. Causal research seeks to find cause and effect relationships between variables. It accomplishes this goal through laboratory and field experiments.

3. Identify data types and sources

There are two types of information available to a market researcher: primary data and secondary data. The next step is to determine the sources of data to be used. The researcher has to decide whether to go for primary data or secondary data. Primary data is original information gathered for a specific purpose. Secondary data refers to information that already exists somewhere and has been collected for some other purpose. Both types of research have a number of activities and methods of conducting associated with them. Secondary research is usually faster and less expensive to obtain than primary research. Gathering secondary research may be as simple as making a trip to a local library or business information center or browsing the Internet. There is already a lot of statistics about different businesses that can be used for this research. After determining which type(s) of information are needed, the methods of accessing data must be determined. There are several different methods of collecting data. These methods include telephone surveys, mail surveys, personal interviews or group surveys. The actual design of the research instrument, the data collection form that is used to ask and record the information is critical to the success of the project. There are two basic methods to collect information: by asking questions or by observing. The most common research instrument is

the questionnaire. There are two types of forms: structured and unstructured. Structured questionnaires list close-end questions. These include multiple choice questions which offer respondents the ability to answer "yes" or "no" or choose from a list of several answer choices. Close-end questions also include scales refer to questions that ask respondents to rank their answers at a particular point on a scale. Unstructured questionnaires have open-ended questions. Respondents can answer in their own words.

Sometimes a combination of both is used. Before going through the time and expense of collecting primary data, one should check for secondary data that previously may have been collected for other purposes but that can be used in the immediate study. Secondary data may be internal to the firm, such as sales invoices and warranty cards, or may be external to the firm such as published data or commercially available data. The government census is a valuable source of secondary data. Secondary data has the advantage of saving time and reducing data gathering costs. The disadvantages are that the data may not fit the problem perfectly and that the accuracy may be more difficult to verify for secondary data than for primary data. Many a time the secondary data might have to be supplemented by primary data originated specifically for the study at hand. Some common types of primary data are:

- Demographic and socioeconomic characteristics
- Psychological and lifestyle characteristics
- Attitudes and opinions
- Awareness and knowledge - for example, brand awareness
- Intentions - for example, purchase intentions. While useful, intentions are not a reliable indication of actual future behavior.
- Motivation - a person's motives are more stable than his/her behavior, so motive is a better predictor of future behavior than is past behavior.

NOTES

Primary data can be obtained by communication or by observation. Communication involves questioning respondents either verbally or in writing. This method is versatile, since one needs to only ask for the information; however, the response may not be accurate. Communication usually is quicker and cheaper than observation. Observation involves the recording of actions and is performed by either a person or some mechanical or electronic device. Observation is less versatile than communication since some attributes of a person may not be readily observable, such as attitudes, awareness, knowledge, intentions, and motivation. Observation also might take longer since observers may have to wait for appropriate events to occur, though observation using scanner data might be quicker and more cost effective. Observation typically is more accurate than communication. Personal interviews have an interviewer bias that mail-in questionnaires do not have. For example, in a personal interview the respondent's perception of the interviewer may affect the responses.

4. Design data collection forms

Once it has been decided to obtain primary data, the mode of collection needs to be decided. Two methods are available for data collection:

1. Observational methods
2. Survey methods

Observational methods: As the name itself suggests, the data are collected through observation. An observer observes and records the data faithfully and accurately. This may be suitable in case of some studies but is not useful to observe attitudes, opinions, motivations and other intangible states of mind. Also in this method, the data collected is non-reactive, as it does not involve the respondent.

Surveys methods: It is one of the most common methods of collecting data for primary marketing research. Surveys can be:

Check your Progress

2. List out the common methods of collecting primary data in marketing research process?

• Personal: The information is sought through personal interviews. A questionnaire is prepared and administered to the respondent during the interview. This is a detailed method of collecting information.

• Telephonic: This is suitable if limited information is sought in a fixed time frame.

• Mail: Here, the questionnaire is sent out in mail and the response is sought. Timely response cannot be sought in this method as there is no control over the survey. All the people to whom the mail was sent may not respond.

Sometimes a combination of two or more methods may be used. Whatever be the method, a structured questionnaire is required to be used. The questionnaire is an important tool for gathering primary data. Poorly constructed questions can result in large errors and invalidate the research data, so significant effort should be put into the questionnaire design. The questionnaire should be tested thoroughly prior to conducting the survey.

5. Determine sampling design and size

A sampling plan is a very important part of the research process. The marketing researcher has to decide whether it will be a sample survey or a census. Definitely a sample survey has its distinct merits. The population from which the sample has to be drawn has to be well defined. A broad choice is to be made between probability sampling and non-probability sampling. The sample design is then chosen depending on the suitability and the availability of the sample frame. The size of the sample chosen is based on statistical methods. This is well defined and also reproduces the characteristics of the population. In practice, however, this objective is never completely attained on account of the occurrence of two types of errors – errors due to bias in the selection and sapling errors.

6. Collect the data

The next step is to collect the data for which the research process has been spelled out. The interviewing and the supervision of field work should be looked into. One of the most difficult tasks is interviewing for

Check your Progress

3. What is a census survey?

marketing research. Many a time the respondents may not part with crucial information unless approached with tact and intelligence. Supervision of field work is important to ensure timely and proper completion of the field survey. If this is not carried out properly, then there results an interview error which may be detrimental to marketing research.

7. Analyze and interpret the data

The next step is to analyze the data that has been collected from the field survey. The raw data is transformed into the right format. First, it is edited so that errors can be corrected or omitted. The data is then coded; this procedure converts the edited raw data into numbers or symbols. A codebook is created to document how the data is coded. Finally, the data is tabulated to count the number of samples falling into various categories. Simple tabulations count the occurrences of each variable independently of the other variables. Cross tabulations, also known as contingency tables or cross tabs, treats two or more variables simultaneously. Cross tabulation is the most commonly utilized data analysis method in marketing research. Many studies take the analysis no further than cross tabulation. Once the tabulation is done, the following analysis can be carried out.

Check your Progress

4. What are the steps in analysis of data?

- **Conjoint Analysis**: The conjoint analysis is a powerful technique for determining consumer preferences for product attributes.
- **Hypothesis Testing**: The null hypothesis in an experiment is the hypothesis that the independent variable has no effect on the dependent variable. The null hypothesis is expressed as H_0. This hypothesis is assumed to be true unless proven otherwise. The alternative to the null hypothesis is the hypothesis that the independent variable does have an effect on the dependent variable. This hypothesis is known as the alternative, research, or experimental hypothesis and is expressed as H_1. Once analysis is completed, make the marketing research conclusion. In order to analyze whether research results are statistically significant or simply by chance, a test of statistical significance can be run.

8. Prepare the research report

Preparing a research report involves three steps:

1. Understanding: Understand the objective of the research and then provide the solution.

2. Organizing: Organize your thoughts and findings and build a logical flow.

3. Writing: Now, draft the outline of your report and then write the same.

It is not sufficient if the market researcher has collected information on the research problem, he also has to interpret the data and draw specific conclusions from it. The results of marketing research must be effectively communicated to management. This report has to be clear and concise.

Types of Report:

Reports can be broadly classified into two types: **oral and written.** Presenting the results of a marketing research study to management generally involves a formal written report as well as an oral presentation. The report and presentation are extremely important. This is because:

1. The results of marketing research are often intangible (there is very little physical evidence of the resources, such as time and effort, that went into the project); the written report is usually the only documentation of the project.

2. The written report and the oral presentation are the only aspect of the study that marketing executives are exposed to, and consequently the overall evaluation of the research project rests on how well this information is communicated. They might not have been part of the marketing research study and hence will not know what information was collected.

3. Since the written research report and oral presentation are typically the responsibility of the marketing research supplier, the communication effectiveness and usefulness of the information provided plays a crucial role in determining the choice of the particular marketing research supplier for the future.

NOTES

Check your Progress

5. What are major types of reports?

Differences between oral and written reports:

An oral report is any presentation that is verbally done to the management. Written reports are documentation of the research findings. There are three major differences between oral and written reports.

1. Oral reports are difficult to interpret as they lack visual advantage. Charts, diagrams or pictures cannot be used to stress on important points. They have to rely only on pauses and volume emphasis.

2. The pace of the presentation cannot be controlled and regulated by the audience who is being presented with the oral report. In a written report, the reader can clarify a certain point by reading it two or more times, if needed slowly and carefully.

3. A researcher will write the report very precisely and with more accuracy since he is aware that a written report is bound to receive considerable attention and scrutiny from the readers. In contrast, an oral report will not be so precise nor will the researcher give as much time in its presentation since it cannot be subjected to the same degree of scrutiny as a written report.

Considerations for Oral Reporting

A researcher has to consider the following when he has been asked to make an oral presentation:

- Consider the audience to whom the report has to be presented and prepare the same so as to incorporate the technical requirements.
- It should be properly planned. The researcher has to rehearse what he is going to say or recommend. He has to collect and organize the data in a logical manner.
- Suitability of the language is another point worth considering. The reporting has to be simple, concise and clear.
- Visual aids like charts, graphs and handouts can be used judiciously if required to create a better impact. However this also needs to be used sparingly as it might disturb the proceedings.

Consideration for Written Reporting

Every person has a different style of writing. There is not really one right style for a report, but there are some basic principles for writing a research report clearly.

Types of Report:

There are many classifications available for reports. Some of them are listed below:

Sl. No.	Classification	Types
1	Time interval	Daily, Weekly, Fortnightly, Monthly or Annual
2	Functional Basis	Informational, Examinational and Analytical Economics, Finance, Industry, and other
3	Subject Matter	subjects of interest
4	Physical-form	Short-form and Long-form
5	Relationship between the reader and the writer	Administrative, Professional and Independent
6	Employment status of authors	Private and Public reports
7	Formality	Formal and Informal

None of these are mutually exclusive. A report may be a combination of some of the above.

Preparation of the Report

Preparing a research report is just not writing alone, but involves other activities besides writing. In fact, writing is actually the last step in the preparation process. Before writing can take place, the results of the research project must be fully understood and thought must be given to what the report will say. The objectives of the research should be matched with the results and the report has to be written.

The general guidelines that should be followed for any report or research paper are as follows:

- **Consider the audience**: The information resulting from the study is going to be used by the marketing managers, who will use the results to

make decisions. Thus, the report has to be understood by them; the report should not be too technical and not too much jargon should be used. This is a particular difficulty when reporting the results of statistical analysis where there is a high probability that few of the target audience have a grasp of statistical concepts. Hence, this needs to be translated into simple language.

• **Be concise, but precise**: Many a time, the researcher in order to convey his effort, tends to overcrowd the report with data. This leads to loss of focus. On the one hand, a written report should be complete in the sense that it stands by itself and that no additional clarification is needed. On the other hand, the report must be concise and must focus on the critical elements of the project and must exclude unimportant issues. Hence, a research report has to be concise and precise.

• **Understand the results and draw conclusions**: The managers who read the report are expecting to see interpretive conclusions in the report. The researcher must therefore understand the results and be able to interpret these. He should analyze the results and present them to the managers.

All the research findings have to be compiled in a report to be then presented to the organization. The format of the marketing research report varies with the needs of the organization. The following outline is the suggested format for writing the research report:

1. Title page
2. Letter of Authorization
3. Summary of findings
4. Table of contents
   - List of tables
   - List of figures
5. Introduction
   - Background to the research problem
   - Objectives
   - Hypotheses
6. Methodology

- Data collection
- Sample and sampling method
- Statistical or qualitative methods used for data analysis
- Sample description

7. Findings
8. Limitations
9. Results, interpretation and conclusions.
10. Recommendations
11. Appendices
12. Bibliography

The report has to be written with objectivity, coherence, clarity in the presentation of the ideas and use of charts and diagrams. Sometimes, the study might also throw up one or more areas where further investigation is required.

2.3 ERRORS IN RESEARCH PROCESS

There are two types of errors

a. Sampling error

b. Non sampling error

Sampling error is the difference between sample value and corresponding population value. Non sampling errors are errors in the difference stages of research as

a. Defective problem design

b. Defective population definition

c. Frame se

d. Surrogate information error

e. Non response error

f. Measurement error

g. Experimental error

h. Poor questionnaire design

i. Interview bias

j. Data processing error

k. Data analysis errors

l. Interpretation errors

2.4 SUMMARY

Marketing Research reduces the uncertainty in the decision-making process and increase the probability and magnitude of success if conducted in a systematic, analytical, and objective manner. The Marketing Research Process involves a number of inter-related activities which have bearing on each other. Each and every step plays an important role in the research process.

2.5 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The steps in marketing research are:

1	Define the research problem
2	Determine research design
3	Identify data types and sources
4	Design data collection forms
5	Determine sampling design and size
6	Collect the data
7	Analyze and interpret the data
8	Prepare the research report

2. The three most common methods of collecting primary data in marketing research are

1. Personal interview

2. Telephone interview

3. Mail questionnaire

3. A census survey would collect data from all the units of the population.

4. The steps in data processing are editing, coding, classification tabulation and processing.

5. The two major types of reports are 1. Oral and 2. Written.

2.6 EXERCISES AND QUESTIONS

1. List out the various steps involved in the Marketing Research Process.

2. It is very important to define the research problem, explain.
3. Classify research designs and explain the relevance of each.
4. What are the types of data sources?
5. Enumerate the methods available for data collection
6. Is it important to determine the sample size? Explain.
7. How will you analyze the data collected from the research?
8. How will you prepare a research report?
9. The various steps in the Marketing Research Process are inter-related. Explain.
10. Does the Market Research Process serve as a framework for finding solution to the research problem at hand? Evaluate critically.

2.7 FURTHER READINGS

1 Boyd Harper W., Ralph Westfall and Stanley F Stasch, *Marketing Research Text and Cases*, Irwin, Inc., 2007.

2 Green, Paul E, and Donald S. Tull, *Research For Marketing Decisions*, Englewood clifs N.J, Prentice Hall, Inc, 2006

3 G.C.Beri, *Marketing Research*, TMH, 2008.

4 Nargundkar R, *Marketing Research Text and Cases,* TMH, 2002

5 Dr.D.D.Sharma, *Marketing research, Principles, applications, and cases,* Sultan Chand and sons, 2007.

UNIT III

RESEARCH DESIGN

STRUCTURE

3.1 Introduction
3.2 Research design
3.3 Concepts relating to research design
3.4 Types of research designs
3.5 Exploratory research studies
3.6 Descriptive research studies
3.7 Causal research studies
3.8 Summary
3.9 Answers to 'Check Your Progress'
3.10 Exercises and Questions
3.11 Further Readings

UNIT OBJECTIVES

- To understand the meaning of Research Design
- To study about the various types of Research designs.
- To understand the type of research design to use for specific problems.

3.1 INTRODUCTION

Research design lays down the methodology involved in the collection of information and arriving at meaningful conclusions from the same. There are many methods for studying and tackling a problem, but there are no perfect methods. Many times more than one method could be used in the research process.

3.2 RESEARCH DESIGN

A research design is the detailed blueprint used to guide a research study toward its objectives. The process of designing a research study involves many interrelated decisions. The most significant decision is the choice of research approach, because it determines how the information will be obtained.

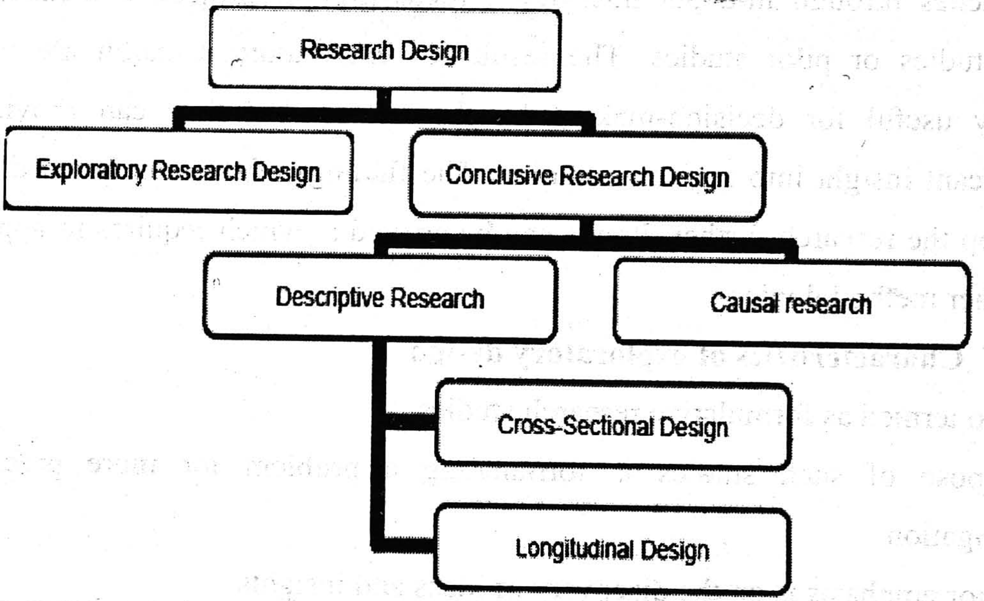
NOTES

According to Green and Tull: A Research Design is the specification of methods and procedures for acquiring the information needed. It is the over-all operational pattern or framework of the project that stipulates what information is to be collected from which sources by what procedures.

3. 3 TYPES OF RESEARCH DESIGNS

There are many classifications accepted for a Research Design. One of the most accepted classification is grouping it under three types:

1. Exploratory
2. Descriptive and
3. Causal



3.4 EXPLORATORY RESEARCH STUDIES

As the term suggests, exploratory research is often conducted because a problem has not been clearly defined as yet, or its real scope is as yet unclear. It is a process of discovery wherein you uncover as many ideas as possible. It allows the researcher to familiarize him/herself with the problem or concept to be studied, and perhaps generate hypothesis to be tested. It expands knowledge. It is the initial research, before more conclusive research is undertaken. Exploratory research helps determine the best research design, data collection method and selection of subjects.

Check your Progress

1. What is an experimental group?

Another common reason for conducting exploratory research is to test concepts before they are launched in the marketplace, always a very costly endeavor. In concept testing, consumers are provided either with a written concept or a prototype for a new, revised or repositioned product, service or strategy. Exploratory research relies more on secondary data. It does not have a rigid design as the researcher themselves are not very well versed with the subject and are trying to gain knowledge of the same. Hence it can be quite informal, relying on secondary research such as reviewing available literature and/or data, or qualitative approaches such as informal discussions with consumers, employees, management or competitors, and more formal approaches through in-depth interviews, focus groups, projective methods, case studies or pilot studies. The results of exploratory research are not usually useful for decision-making by themselves, but they can provide significant insight into a given situation. The findings of this can be used to develop the research further. Points can be arrived at which requires to apply the other methodologies.

3.4.1 Characteristics of exploratory design

1. Also termed as formulative research studies.

2. Purpose of such studies is formulating a problem for more precise investigation.

3. Major emphasis is on the discovery of ideas and insights.

4. Research design has to be flexible enough to provide opportunity for considering different aspects of a problem under study.

5. Inbuilt flexibility is essential.

3.4.2 Three methods of conducting exploratory research

Check your Progress

2. What is a control group?

Following are three methods in the context of research design for studies:

1. The survey of concerning literature

2. The experience survey

3. The analysis of insight –stimulating examples.

NOTES

3.4.3 The survey of concerning literature:

This happens to be the most simple and fruitful method of formulating the research problem. Hypothesis stated by earlier workers may be reviewed and their usefulness be evaluated as a basis for further research. In this way researcher should review and build upon the work already done by others, but in cases where hypothesis has not been formulated hi task is to review the available material for deriving the relevant hypothesis from it.

3.4.4 Experience Survey:

It is the survey of people who have had practical experience with the survey to be studied. The object is to obtain insight into the relationship between variables and new ideas relating to the research problem. For such a survey people who are competent and can contribute new ideas may be carefully selected as respondents to ensure representation of different of experience. The respondents selected can be interviewed by the investigator. An interview schedule is prepared by the researcher for systematic questioning of informants. The interview must ensure flexibility in the sense that the respondents should be allowed to raise issues and questions which the investigator has not previously considered. The interview may last for few hours. Hence, it its often considered desirable to send a copy of the questions to be discussed to the respondents well in advance. This gives an opportunity to the respondents for doing some advance thinking over various issues involved so that, at the time of interview they may be able to contribute effectively. Thus, an experience survey may enable the researcher to define the problem more concisely ana help in formulation of research hypothesis. This survey may as well provide information about the practical possibilities for doing different types of research.

Check your Progress

3. What is a treatment?

3.4.5 Analysis of insight stimulating examples:

This is a fruitful method for suggesting hypothesis for research. It is particularly suitable in areas where there is little experience to serve as a guide. It consists of the intensive study of the selected instances of the phenomenon in which on is interested. For this purpose the existing

records may be examined the unstructured interviewing may take place or some other approach may be adopted. Attitude of the investigator, the intensity of the study and the ability of the researcher to draw together diverse information into a unified interpretation are the main features which make this method an appropriate procedure for evoking insights. Examples for the above are:

- Reactions of strangers
- Reactions of marginal individuals
- Study of individuals who are in a transition from one stage to another.
- Reactions of individuals from different social strata.

3.5 DESCRIPTIVE AND DIAGNOSTIC RESEARCH STUDIES

Descriptive research studies are concerned with describing the characteristics of certain individuals or a group. E.g. studies concerning whether certain variables are associated. Diagnostic research studies determine the frequency of with which something occurs or its association with something else. E.g. studies concerned with specific predictions, with narration of facts and characteristics concerning individual, group or situation. The descriptive as well as diagnostic research studies share common requirements. In both the studies, the researcher must be able to define clearly, what he wants to measure and must find adequate methods of measuring it. The aim is to obtain complete and accurate information, hence, the procedure to be used must be carefully planned. It should make enough provision for protection against bias and must maximize reliability. The design must be rigid and not flexible.

Descriptive research or statistical research provides data about the population or universe being studied. It describes the "who, what, when, where and how" of a situation and not what caused it. Therefore, descriptive research is used when the objective is to provide a systematic description that is as factual and accurate as possible. It provides the

number of times something occurs, or frequency, lends itself to statistical calculations such as determining the average number of occurrences or central tendencies. One of its major limitations is that it cannot help determine what causes a specific behaviour, motivation or occurrence. In other words, it cannot establish a causal research relationship between variables. The four most common types of descriptive research designs are

3.5.1. Observation: Observation is a primary method of collecting data by human, mechanical, electrical or electronic means. The researcher may or may not have direct contact or communication with the people whose behaviour is being recorded. Observation techniques can be part of qualitative research as well as quantitative research techniques. The commonly used observation methods are:

· *Participant and non participant observation:* This depends on whether the researcher chooses to be part of the situation s/he is studying. (e.g. studying team dynamics by being a team member would be participant observation)

· *Obtrusive and unobtrusive observation:* Depends on whether the subjects being studied can detect the observation (e.g. hidden microphones or cameras observing behaviour)

· *Observation in natural or contrived settings:* Observing the behaviour in its natural setting and in a condition where the natural settings are created.

· *Disguised and non-disguised observation:* Depends on whether the subjects being observed are aware that they are being studied or not. In disguised observation, the researcher may not disclose his true identity and pretend to be someone else to keep away the bias in the findings.

Check your Progress

4. What is a disguised observation?

· *Structured and unstructured observation:* This refers to guidelines or a checklist being used for the aspects of the behaviour that are to be recorded; for instance, noting who starts the introductory conversation between the group members and what specific words are used by way of introduction.

· *Direct and indirect observation:* This depends on whether the behaviour is being observed during the time it occurs or after the occurrence, as in the case

of TV viewing, for instance, where choice of program and channel flicking can all be recorded for later analysis.

One distinct advantage of the observation technique is that it records actual behaviour, not what people say they said/did or believe they will say/do. On the other hand, the observation technique does not provide us with any insights into what the person may be thinking or what might motivate a given behaviour/comment. This type of information can only be obtained by asking people directly or indirectly.

3.5.2. Surveys: The survey technique mainly involves the collection of primary data about subjects, usually by selecting a representative sample of the population or universe under study, through the use of a questionnaire. It is a very popular since many different types of information can be collected, including attitudinal, motivational, ehavioural and perceptive aspects. It allows for standardization and uniformity in the questions asked and in the method of approaching subjects, making it easier to compare and contrast answers by respondent group. It also ensures higher reliability than some other techniques. If properly designed and implemented, surveys can be an efficient and accurate means of determining information about a given population. Results can be provided relatively quickly, and depending on the sample size and methodology chosen, they are relatively inexpensive. However, surveys also have a number of disadvantages, which must be considered

by the researcher in determining the appropriate data collection technique. Since in any survey, the respondent knows that s/he is being studied, the information provided may not be valid insofar as the respondent may wish to impress (e.g. by attributing him/herself a higher income or education level) or please (e.g. researcher by providing the kind of response s/he believes the researcher is looking for) the researcher. This is known as response error or bias. The willingness or ability to reply can also pose a problem. If the information sought is considered sensitive or intrusive the respondent may hesitate to reply, leading to a high rate of refusal. This can be overcome by framing such questions carefully. There can be an interviewer error or bias as

the interviewer can (inadvertently) influence the response elicited through comments made or by stressing certain words in the question itself. This is seen through facial expressions, body language or even the clothing that is worn. Another consideration is response rate. Depending on the method chosen, the length of the questionnaire, the type and/or motivation of the respondent, the type of questions and/or subject matter, the time of day or place, and whether respondents were informed to expect the survey or offered an incentive can all influence the response rate obtained. Proper questionnaire design and question wording can help increase response rate.

3.5.4 Case Method Design

This method is not often used in descriptive research. It is more widely used in exploratory research. The procedure is the same as in exploratory research. The only difference between the two lies in the fact that while exploratory research offers flexibility, descriptive research is more structured, clearly defining the research problem and the points to be investigated. Researchers use analogy as a method of analysis in cases. Through cases, researchers attempt to find

1. Common features to all cases in a general group
2. Features which are not common to all groups, but common to some subgroups
3. Features unique to a specific case
4. Features that are common and those that are uncommon are analysed thoroughly to formulate hypothesis. Researchers should be careful in selecting the cases for analysis.

3.5.4.1 Advantages of Case Method

Advantages are:

1.Cases are studied comprehensively, taking into consideration all aspects.

2.Unlike statistical studies which involve abstracts from real situation, case study describes a real-time situation.

3.5.4.2 Disadvantages of Case Method

Some of the disadvantages of case study are :

1. Case analysis is very subjective.
2. It is difficult to have a formal research method.
3. Researchers tend to generalize the situations, though the case may not call for such generalizations.

3.5.4 Statistical studies

This is the most widely used method in Marketing Research. It makes use of techniques that vary from simple means and percentages to very sophisticated techniques. In this method, a limited number of factors from a large number of cases are studied in-depth.

3.5.4.1 Use of statistical method

Statistical tools are used by most marketing research professionals to understand the dynamics of the market. Data is usually collected through observation or through interviewing. Surveys to find the consumption patterns of children ad generate profiles of the people using the internet to make purchases are examples of the statistical method. Let us see how Forrester research, the leading global marketing research agency, has studied online purchases. They have discovered that the average online buyer reports purchases in two product categories : books and software.

The following are some of the observations made by Forrester on line purchases.

- Media and technology lead net shopping. More than 50% of online shoppers buy software and more than 40% buy books. Buyers spend an average of more than $100 over three- month period.
- More than 250.000 North Americans bought cars and computers online.
- Three percent of online consumers report online securities trades.

3.5.4.2 Advantages of Statistical Method

Statistical study involves a large number of interviews or observation. Statistical techniques are used specifically for mass data. Two different

researchers conducting statistical research will arrive at the same results, while two people using the case study method may not arrive at the same results. This is mainly because of the subjectivity of the case study method. Statistical study helps the researcher to make more accurate generalizations. If the sampling is properly done, the generalization will be universally true.

Difference between exploratory and descriptive research

Research Design	Types of study	
	Exploratory / formulative	*Descriptive* / Diagnostic
Overall design	Flexible design (design must provide opportunity for considering different aspects of the problem)	Rigid design (design must make enough provision for protection against and must maximize reliability)
sampling design	Non- probability sampling design (purposive or judgement sampling)	Probability sampling design (random sampling)
statistical design	No pre-planned design for analysis	Pre-planned design for analysis
observational design	Unstructured instruments for collection of data	Structured or well thought out instruments for collection of data
operational design	No fixed design about the operational procedure	Advanced decisions about operational procedures

3.5.4.3 Disadvantages

· Fails to prove cause-and-effect relationship

· Direction of causal effect is not clear in statistical studies.

Statistical Descriptive studies are also classified into:

1. Cross-sectional studies: It deals with a sample of elements from a given population. Number of characteristics from the sample elements are collected and analyzed. It is of two types: field studies and surveys. Cross-sectional studies sample the population to make measurements at a specific point in time. A special type of cross sectional analysis is a cohort analysis, which tracks an aggregate of individuals who experience the same event within the same time interval over time. You can use Cohort analyses for long forecasting of product demand.

NOTES

2. *Longitudinal studies:* This is based on panel data and panel methods. A panel constitutes a group of respondents who are interviewed and re- interviewed from time to time. Hence the same variable is repeatedly measured. This helps in studying a particular behaviour over a period of time. Longitudinal studies are time series analyses that make repeated measurements of the same individuals, thus allowing you to monitor behaviour such as brand switching. However, longitudinal studies are not necessarily representative since many people may refuse to participate because of the commitment required.

3.6 CAUSAL RESEARCH

Causal research is undertaken to see if there is a cause and effect relationship between variables. In order to determine causality, it is important to hold the variable that is assumed to cause the change in the other variable(s) constant and then measure the changes in the other variable(s). This type of research is very complex and the researcher can never be completely certain that there are not other factors influencing the causal relationship, especially when dealing with people's attitudes and motivations. There are often much deeper psychological considerations that even the respondent may not be aware of. There are two research methods for exploring the cause and effect relationship between variables:

Check your Progress

5. What is a longitudinal study?

1. **Experimentation or natural experimentation:** This highly controlled method allows the researcher to manipulate a specific independent variable in order to determine what effect this manipulation would have on other dependent variables. Experimentation also calls for a control group as well as an experimentation group, and subjects would be assigned randomly to either group. The researcher can further decide whether the experiment should take place in a laboratory or in the field, i.e. the "natural" setting as opposed to an "artificial" one. Laboratory research allows the researcher to control and/or eliminate as many intervening variables as possible.

Simulation: Another way of establishing causality between variables is through the use of simulation. A sophisticated set of mathematical formula are

used to simulate or imitate a real life situation. By changing one variable in the equation, it is possible to determine the effect on the other variables in the equation. For the natural experiments there are three classes of designs:

1. Time-series and trend designs
2. Cross-sectional designs and
3. A combination of the above two.

Time series and trend designs: In a time series design, data is collected from the sample or population at successive intervals. The trend data relate to matched samples drawn from the same population at successive intervals. It can be of many types. A simple design can be represented as below:

$$X \qquad O$$

Where X indicates the exposure of a group to an experimental treatment and O indicates the observation or measurement taken on the subject or group after an experimental treatment. Another method also involves a control group. This can be represented as below:

$$O_1 \quad O_2 \quad O_3 \quad X \quad O_4 \quad O_5 \quad O_6$$
$$O'_1 \quad O'_2 \quad O'_3 \qquad O'_4 \quad O'_5 \quad O'_6$$

Where O's represent measurement of the control group. This is termed as multiple time-series design.

Cross-sectional designs: It studies the effect of different levels of treatments on several groups at the same time. It can be represented as below:

$$X_1 \quad O_1$$
$$X_2 \quad O_2$$
$$X_3 \quad O_3$$
$$X_4 \quad O_4$$

An example would be different kind of incentives given for the same product in various territories. This would help in understanding the effect of varying the incentive on the sales performance across territories.

Combinational Design: This design combines both the time-series and cross sectional designs. This design is generally seen while measuring advertising effectiveness in a panel. An advertisement is run and the respondents are asked if they have seen it earlier. Those who have seen it earlier constitute the test group and those who have not constitute the control group. The purchase made before and after the advertisement by the test and the control group marks the advertising effectiveness. So many research designs have been listed. The one that is ultimately selected should help in solving the problem. It should help in arriving at the desired conclusions.

3.7 EXPERIMENTAL DESIGN

Experimental studies are those where the researcher tests the hypothesis of casual relationship between variables. Such studies require procedures that will not only reduce bias and increase reliability, but will permit drawing inferences about casuality. Professor R.A. Fisher begun such designs when he was working at Rothamsted Experimental Station (Centre for Agricultural Research in England). Professor Fischer found that by dividing agricultural fields or plots into different blocks and then by conducting experiments in each of these blocks, the information collected and inferences drawn happen to be more reliable. This fact inspired him to develop certain experimental designs for testing hypotheses concerning scientific investigation.

3.7.1 Basic Principles of Experimental Design

Professor Fisher has enumerated three principles of experimental designs:

1. the Principle of Replication;
2. the Principle of Randomization; and the
3. the Principle of Local Control.

According to the *Principle of Replication,* the experiment should be repeated more than once. Thus, each treatment is applied in many experimental units instead of one. By doing so the statistical accuracy of the experiments is increased. The entire experiment can even be repeated several times for better results. Conceptually replication does not present any difficulty, but

computationally it does. It should be remembered that replication is introduced in order to increase the precision of a study; that is to say, to increase the accuracy with which main effects and interactions can be estimated.

The *Principle of Randomization* provides protection, when we conduct an experiment, against the effects of extraneous factors by randomization. In other words, this principle indicates that we should design or plan the experiment in such a way that the variations caused by extraneous factors can be combined under the general heading of "chance".

The *Principle of Local Control* is another important principle of experimental designs. Under it the extraneous factors, the known source of variability, is made to vary deliberately over as wide a range as necessary and this needs to be done in such a way that the variability it causes can be measured and hence eliminated from the experimental error. This means that we should plan the experiment in manner that we can perform a two-way analysis of variance, in which the total variability of the data is divided into three components attributed to treatments (the subject), the extraneous factors and experimental error.

3.7.2 Formal and Informal Experimental Design

Experimental design refers to the framework or structure of the experiment and as such there are several such experimental design. Experimental design can be classified into two broad categories. Informal experimental design and Formal experimental design. Informal experimental design are those design that normally uses a less sophisticated form of analysis based on differences in magnitude, whereas formal experimental design offer relatively more control and use precise statistical procedures for analysis. Important statically designs are as follows:

1. Informal experimental design:

- Before and after without control design.
- After only with control design.
- Before and after with control design.

1.Formal experimental design:

- Completely randomized design (C.R.design).
- Randomized block design (R. B. design).
- Latin square design (L.S. design).
- Factorial design.

The details of each of the above stated formal and informal experimental design are explained as follows.

1.7.2.1 Before-and-after control design:

In such a design a single test group or area is selected and the dependant variable is measured before the introduction of the treatment. The treatment is then introduced and the dependant variable is measured again after the treatment has been introduced. The effect of the treatment would be equal to the level of the phenomena after the treatment minus the level of phenomenon before the treatment. The main difficulty of such a design is that with the passage of the time considerable extraneous variation may be there in the treatment effect. The design can be represented as

Test area:	Level of phenomenon Before treatment (X)	Treatment introduced →	Level of phenomenon after the treatment (Y)

Treatment Effect = (Y) - (X)

1.7.2.2 After only with control design:

In this design two group of arise (test area and control area) are selected and the treatment is introduced in the test area only. The dependant variable is then measured in both the areas at the same time. Treatment impact is assessed by subtracting the value of the dependant variable in the control area from the value in the test area. This can be exhibited in the following form

Test area:	Treatment introduced	Level of phenomenon after treatment (Y)
Control area	→	Level of phenomenon without treatment (Z)

Treatment effect = (Y) – (Z)

The basic assumption in such a design is that the two areas are identical with respect to their behavior towards the phenomenon considered. If the

assumption is not true, then there is then there is the possibility of extraneous variant entering into the treatment effect. However, data can be collected in such a design without the introduction of the problems with the passage of time. In this respect this design is superior to before-and-after without control design.

1.7.2.2 Before-and-after with control design:

In this design two areas are selected and the dependent variable is measured in both the areas for identical time period before the treatment .The treatment is then introduced into the test area only, and the dependent area is measured both for an identical time period after the introduction of the treatment. The treatment effect is determined by subtracting change in the dependent variable in the control area from the change in the dependent variable in the test area. This can be shown in the following way:

	Time Period 1		Time Period 2
Test area:	Level of phenomenon before treatment (X)	Treatment introduced →	Level of phenomenon after treatment(Y)
Control area:	Level of phenomenon Without treatment (A)		Level of phenomenon without treatment (B)

Treatment Effect = (Y-X) – (Z-A)

This design is superior to the other two design for the simple reason that it avoids extraneous variation resulting both from the passage of time and from non-comparability of the test and control areas. But at times due to lack of historical data, time, so it is preferred to select one of the first two informal designs stated above.

1.7.2.3 Completely randomized design (C.R.):

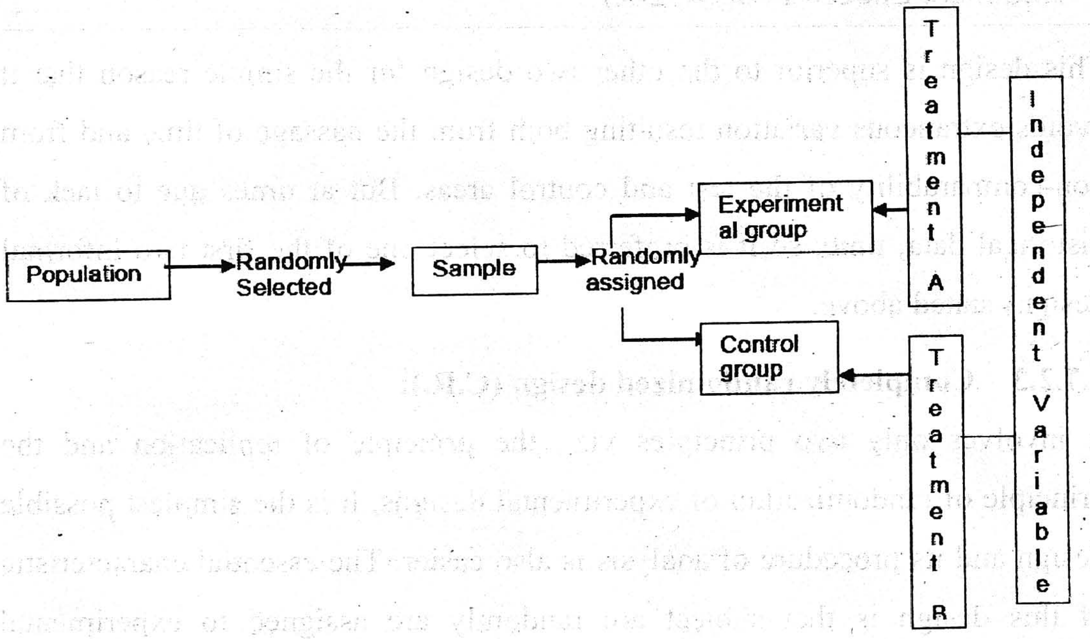
It involves only two principles viz., the principle of replication and the principle of randomization of experimental designs. It is the simplest possible design and its procedure of analysis is also easier. The essential characteristic of this design is that subject are randomly are assigned to experimental treatment. For instance, if we have 10 subject and we to test 5under treatment A and 5 under treatment B, the randomized process gives the every possible group of 5 subjects selected from the group of 10 an equal opportunity of

being assigned to treatment A and treatment B. One-way analysis of variance (or one way ANOVA) is used to analyze such a design. Even unequal application works in this design. It provides maximum number of degree of freedom to the error. Such a design is used when experimental areas happen to be homogenous. Technically, when all the variation due to uncontrolled extraneous factors are included under the heading of chance variation, we refer to the design of experiment as C.R.design. The brief description on the two form of such design is explained below:

3.7.2.4.1 Two-group simple randomized design:

In a two-group simple randomized design, first of all the population is defined and then from the population a sample is selected randomly. Further requirement of this design is that items, after being selected randomly from the population, be randomly assigned to the experimental and control groups(such random assignmentof items of two group is called as principle of randomization.). Thus this design yields two groups as representative of the population. In the diagram form this design can be shown in this way

Two-group simple randomized design in Diagram form



Since in the simple randomized design the elements constituting the sample are randomly from the same population and randomly from the same population and randomly assigned to the experimental and control groups, it

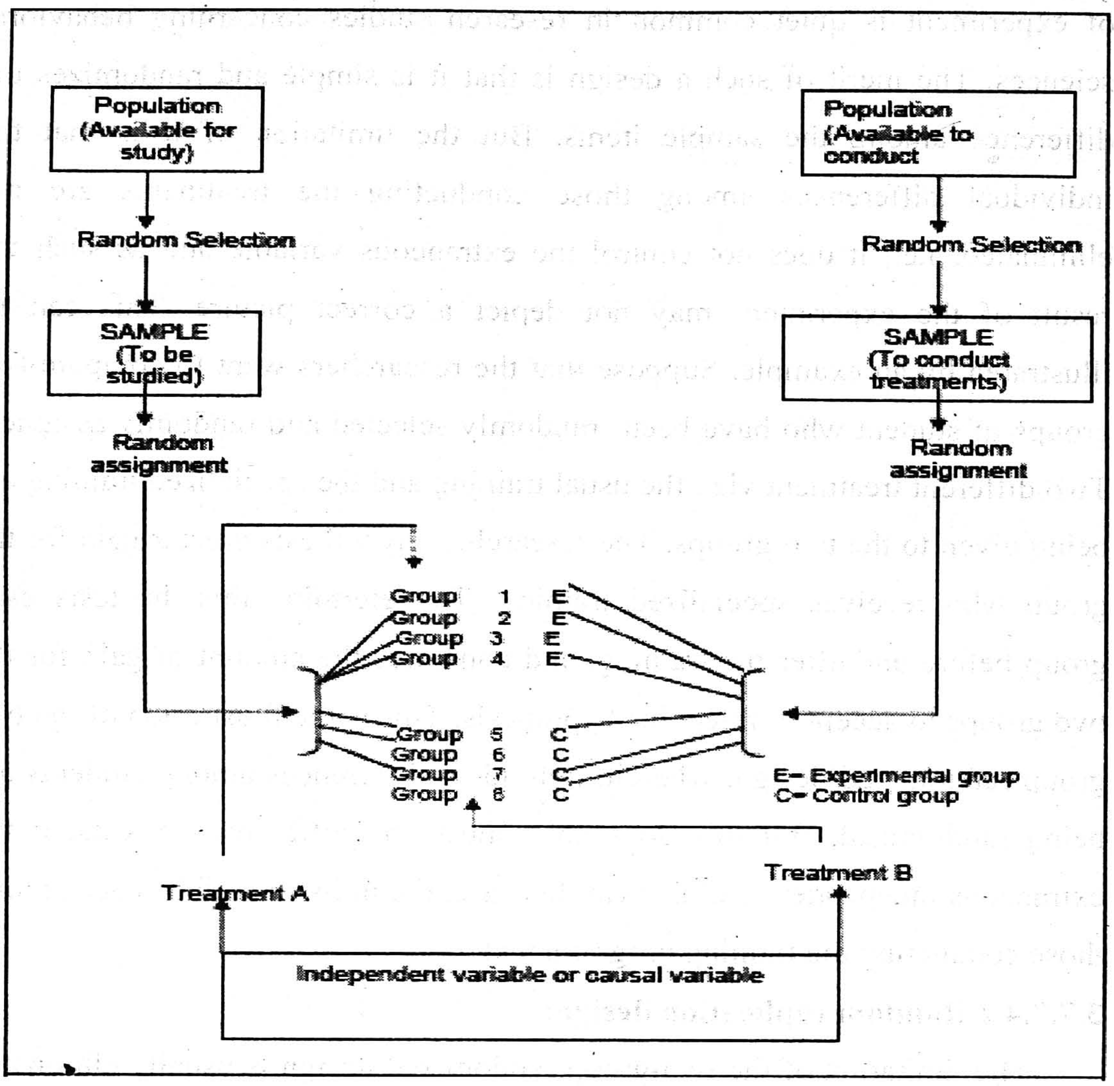
becomes possible to draw conclusion on the basis of samples applicable for the population. The two group (experimental and control groups) of such a design are given different treatment of the independent variable. This design of experiment is quiet common in research studies concerning behavioral sciences. The merit of such a design is that it is simple and randomizes the difference among the sample items. But the limitation of it is that the individual differences among those conducting the treatments are not eliminated, i.e., it does not control the extraneous variable and as such the result of the experiment may not depict a correct picture. This can be illustrated by an example. Suppose that the researchers want to compare two groups of student who have been randomly selected and randomly assigned. Two different treatment viz., the usual training and the specialized training are being given to the two groups. The researchers hypothesis greater gain for the group who receives specialized training. To determine this, he tests each group before and after the training, and compares the amount of gain for the two groups to accept or reject his hypothesis. This is the illustration of the two group randomized design, wherein individual differences among students are being randomized. But this does not control the differential effects of the extraneous independent variable (in this case, the individual difference among those conducting the training programmes)

3.7.2.4.2 Random replication design:

The limitation of the two-group randomized design is usually eliminated within the random replications design. In the illustration just cited above, the *teacher differences* on the dependent variable were ignored, i.e., the extraneous variable was not controlled. But in a random replications design, the effect of such differences are minimised (or reduced) by providing a number of repetitions for each treatment. Each repetition is technically called a 'replication'. Random replication design serves two purposes viz., it provides controls for the differential effects of the extraneous independent variables and secondly, it randomizes any individual differences among those

conducting the treatments. From the diagram it is clear that there are two populations in the replication design.

Random replication design



The sample is taken randomly from the population available for study and is randomly assigned to, say, four experimental and four control groups. Similarly, sample is taken randomly from the population available to conduct experiments (because of eight groups eight such individuals be selected) and the eight individuals so selected should be randomly assigned to the eight groups. Generally, equal number of items are put in each group so that the size of the group is not likely to affect the results of the study. Variables relating to both population characteristics are assumed to be

randomly distributed among the two groups. Thus, this random replication design is, in fact an extension of the two-group simple randomized design.

Randomized block design (R.B design)

It is an improvement over the C.R. design. In the R.B design the principle of local content can be applied along with the other two principles of experimental designs. In the R.B. design, subjects are first divided into groups, known as blocks, such that within each group the subjects are relatively homogeneous in respect to some selected variable. The variable selected for grouping the subjects is one that is believed to be related to the measures to be obtained in respect of the dependent variable. The number of subjects in a given block would be equal to the number of treatments and one subject in each block would be randomly assigned to each treatment. In general, blocks are levels at which we hold the extraneous factor fixed, so that its contribution to the total variability of data can be measured. The main feature of the R.B design is that in this each treatment appears the same number of times in each block. The R.B design is analysed by the two-way analysis of variance (two-way ANOVA) technique. Let us illustrate the R.B. design with the help of an example. Suppose four different forms of a standardized test in statistics were given to each of five students (selected one from each of the five I.Q. blocks) and following are the scores which they obtained.

	Very low IQ	Low IQ	Average IQ	High IQ	Very High IQ
	Student A	**Student B**	**Student C**	**Student D**	**Student E**
Form 1	82	67	57	71	73
Form 2	90	68	54	70	81
Form 3	86	73	51	69	84
Form 4	93	77	60	65	71

If each student separately randomized the order in which he or she took the four tests (by using random numbers or some similar device), we refer to the design of this experiment as a R.B. design. The purpose of this randomization is to take care of such possible extraneous factors (say as fatigue) or perhaps the experience gained from repeatedly taking the test.

3.7.2.6. Latin squares design (L.S design)

It is an experimental design very frequently used in agricultural research. The conditions under which agricultural investigations are carried out are different from those in other studies for nature plays an important role in agriculture. For instance, an experiment has to be made through which the effects of five different varieties of fertilizers on the yield of a certain crop, say wheat, it to be judged. In such a case the varying fertility of the soil in different blocks in which the experiment has to be performed must be taken into consideration; otherwise the results obtained may not be very dependable because the output happens to be the effect not only of fertilizers, but it may also be the effect of fertility of soil. Similarly there may be the impact of varying seeds on the yield. To overcome such difficulties, the L.S design is used when there are two major extraneous factors such as the varying soil fertility and varying seeds. The merit of this experimental design is that it enables differences in fertility gradients in the field to be eliminated in comparison to the effects of different varieties of fertilizers on the yield of the crop. But this design suffers from one limitation, and it is that although each row and each column represents equally all fertilizer varieties, there may be considerable difference in the row and column means both up and across the field. This, in other words, means that in L.S. design we must assume that there is no interaction between treatments and blocking factors.

3.7.2.7. Factorial designs:

Factorial designs are used in experiments where the effects of varying more than one factor are to be determined. They are specially important in several economic and social phenomena where usually a large number of factors affect a particular problem. Factorial designs can be of two types: (i) simple factorial designs and (ii) complex factorial designs.

3.7.2.7.1 S*imple factorial designs*: In case of simple factorial designs, we consider the effects of varying two factors on the dependent variable, but when an experiment is done with more than two factors, we use complex

factorial designs. Simple factorial design is also termed as a 'two-factor-factorial design,' whereas complex factorial design is known as 'multi-factor-factorial design.' Simple factorial design may either be a 2 x 2 simple factorial design, or it may be, say 3 x 4 or 5 x 3 or the like type of simple factorial design. Illustration : (4* 3 simple factorial design) The 4*3 simplex factorial design will usually include four treatments of the experimental variable and three levels of the control variable. Graphically it may take following form:

	Experimental Variable			
CONTROL VARIABLE	TREATMENT A	TREATMENT B	TREATMENT C	TREATMENT D
Level 1	Cell 1	Cell 4	Cell 7	Cell 10
Level 2	Cell 2	Cell 5	Cell 8	Cell 11
Level 3	Cell 3	Cell 6	Cell 9	Cell 12

This model of a simplex factorial design includes four treatments viz. A, B, C and D of the experimental variable and three levels viz (I), (II), and (III) of the control variable and has 12 different cells as shown above. This shows that a 2 * 2 simple factorial design can be generalized to any number of treatments and levels. In such a design the means for the columns provide the researcher with an estimate of the main effects for the levels. Such a design also enables the researcher to determine the interaction between treatments and levels

3.7.2.7.2 Complex factorial designs: experiments with more than two factors at a time involve the use of complex factorial designs. A design which considers three or more independent variables simultaneously is called a complex factorial design.

		Experimental Variable			
		Treatment A		Treatment B	
		Control Variable 2 Level I	Control Variable 2 Level II	Control Variable 2 level I	Control Variable 2 Level II
Control Variable 1	Level I	Cell 1	Cell 3	Cell 5	Cell 7
	Level II	Cell 2	Cell 4	Cell 6	Cell 8

In case of three factors with one experimental variable having two levels, the design used will be termed 2 * 2 * 2 * complex factorial design which will contain a total of eight cells as shown above.To obtain the first order interaction say, for EV * CV 1in the above stated design, the researcher must necessarily ignore control variable 2 for which purpose he may develop 2 * 2 design from the 2 * 2* 2 design by combining the data of the relevant cells of the latter design as has been shown below:

		Experimental Variable	
		Treatment A	Treatment B
Control	Level 1	Cells 1, 3	Cells 5, 7
	Level 2	Cells 2, 4	Cells 6, 8

Similarly, the researcher can determine other first order interactions analysis of the first order interaction, here essentially a simple factorial analysis as only two variables are considered at a time and the remaining on eis ignored. But the analysis of the second interaction would not ignore one of the three independent variables in the case of a 2 * 2* 2 design. The analysis would be termed as a complex factorial analysis. Factorial analysis are used mainly because of two advantages :

1. They provide equivalent accuracy (as happens in the case of experiments with only one factor) with less labour and as much are a source of economy. Using factorial designs, we can determine the main effects of two (in simple factorial design) or more (in case of complex factorial design) factors (in simple factorial design) or more (in case of complex factorial design) factors (or variables in one single experiment.
2. They permit various other comparisons of interest. For example, they give information about such effects which cannot be obtained by treating one single factor at a time. The determination of interaction effects is possible in case of factorial design.

There are several research designs and the researcher must decide in advance of collection and analysis of data as to which design which prove to be more appropriate for his research project. One must give due weight to

various points such as type of universe and its nature, objective of his study, source list or sampling frame desired standard of accuracy and the like when taking decision in respect of the design for his research project.

3.8 SUMMARY:

A Research Design is the specification of methods and procedures for acquiring the information needed. It is the blueprint for a research process. There are many classifications accepted for a Research Design.

3.9 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A group to exposed to some special conditions it is termed as experimental group
2. A group is exposed to usual conditions; it is termed a control group.
3. The different conditions under which experimental and control groups are put are usually referred to as 'treatments'.
4. In a Disguised observation, the researcher may not disclose his true identity and pretend to be someone else to keep away the bias in the findings.
5. Longitudinal study: This is based on panel data and panel methods. A panel constitutes a group of respondents who are interviewed and re-interviewed from time to time. Hence the same variable is repeatedly measured. This helps in studying a particular behaviour over a period of time. Longitudinal studies are time series analyses that make repeated measurements of the same individuals, thus allowing you to monitor behavior such as brand switching.

3.10 QUESTIONS

1. What is a Research Design? Explain.
2. Classify the Research Designs and define each of them.
4. 'Conclusive research helps in drawing conclusions'. Explain.
5. What is descriptive Research? Classify and explain the same.
6. What are cross-sectional and longitudinal studies? Explain.
7. What is causal research? Explain the causal research methods.
8. What is Time series design? When is it used?

NOTES

3.11 FURTHER READING

1 Boyd Harper W., Ralph Westfall and Stanley F Stasch, *Marketing Research Text and Cases*, Irwin, Inc., 2007.

2 Green, Paul E, and Donald S. Tull, *Research For Marketing Decisions*, Englewood clifs N.J, Prentice Hall, Inc, 2006

3 G.C.Beri, *Marketing Research*, TMH, 2008.

UNIT IV

BASIC METHODS OF COLLECTING DATA

STRUCTURE

4.1 Introduction

4.2 Secondary data

4.3 Primary data

4.4 Observation

4.5 Surveys

4.6 Case study

4.7 Personal Interviews

4.8 Data collection instruments

4.9 Errors in data collection

4.10 Difficulties in data collection

4.11 Summary

4.12 Answers to 'Check Your Progress'

4.13 Exercises and Questions

4.14 Further Readings

UNIT OBJECTIVE

After learning this unit you would be able to

- To understand the meaning and importance of data sources.
- To read in detail about the sources of Primary data and secondary data sources.
- To understand the relevance of these data sources while solving a research problem.

4.1 INTRODUCTION

One of the most important components of Marketing Research is collection of data required to solve a defined research problem.

The most important point to be considered before this is to research the secondary sources and gather data already available. This gives a logical perspective to problem solving. Only then the actual data required to be collected from the primary survey can be well defined. so it is imperative to know the advantages and drawbacks of Secondary and Primary data.

4.2 SECONDARY DATA

Secondary data is defined as the data that has been collected by individuals or agencies for purposes other than those of the particular research study. For example, if a government department has conducted a survey of, say, school going children, then a uniform manufacturer might use this data for his research purpose. As mentioned earlier, it is ideal to undertake a marketing research study after a prior search of secondary sources (also termed desk research).

4.2.1 Disadvantages of secondary data:

Even though the secondary data offers a lot of advantages; it also has its own shortcomings. This corresponds to both the source and the quality of the data. The main disadvantages may be listed as follows:

1. The researcher has to be careful while using the units defined in the data. It is better to study the definitions used prior to accepting the same for research purpose.
2. The errors of measurements are not generally published in secondary sources and hence this should be considered while looking at data from secondary sources.
3. The reliability of published data may vary over time. Hence the data needs to be checked for time validity.
4. Many a time the data collected may be outdated and hence it needs to be refreshed again. This may otherwise hinder the analysis.

4.2.2 Secondary sources of information:

Secondary sources of information can be collected from two sources: internal sources and external sources.

4.2.2.1 Internal sources of secondary information:

Lot of data is available within an organization regarding day to day operations. These data can be utilized wherever required. These include:

a. Sales data: Sales orders are received, invoiced and delivered. Cost of the goods supplied is also recorded. Sales across different territories are recorded via the reports received from the field. Most of these reports can be used for making marketing decisions. These resources are generally overlooked while deciding on critical issues. Lot of information pertaining to sales by territory, sales by customer type, prices and discounts, average size of order by customer, customer type, geographical area, average sales by sales person and sales by pack size and pack type, etc. This data can be used to identify the most profitable product and customers, tracking sales trends, analysis on discounts given, scattering pattern of sales orders, effect of seasonality on sales, etc.

b. Financial data: This relates to data on various costs involved in procurement of raw materials, production of goods, distribution of goods, conversion costs, labor costs, transportation cost, storage cost, etc. With such data the efficiency of operation can be determined. It helps in assessing the cost of production of a new product, analysis the cost of free capacity, etc.

c. Transportation data: A good record of the data relating to transport operations determine which route to use, which transporter to use, cost of effective routing patterns, etc. This helps in determining whether it would be sensible to have your

own vehicle or hire a vehicle. This enables decision towards a trade off analysis towards a better profitability.

d. Human Resource data: Enormous information could be collected from an organizational perspective from the human resource department. Data on employee turnover, absenteeism, strength of employees could be obtained. This would help in man power planning for the present and future, succession planning, training and development for better productivity, etc.

e. Storage data: This may help in calculating the direct product profitability by calculating the rate of stock turn; stock handling costs, assessing the efficiency of certain marketing operations and the efficiency of the marketing system as a whole. External sources of secondary information

4.2.2.2 External secondary sources

The main sources of external secondary sources are (1) government (federal, state and local) (2) trade associations (3) commercial services (4) national and international institutions.

1. Government statistics: These may include all or some of the following:

Population census, Social surveys, family expenditure surveys, Import/export statistics, Production statistics, and Agricultural statistics. Some of the Indian government bodies are:

· Population Statistics of Govt. of India – Provides statistics related to general population of India

· Central Bureau of Health Intelligence – Provides health related statistics

· Indian Council of medical Research – Provides information on research being conducted on major diseases

· Policy Reform Options Database – Provides data on policy reforms

Ministry of Health and Family Welfare – Provides information on family welfare

· Ministry of Statistics and Programme implementation – Gives information on various statistical indicators of Indian economy

· India Brand Equity Foundation – Provides information on Indian economy and Industry

· Insurance Regulatory and Development Authority – Data on Health Insurance in India.

2. Trade associations - They might produce a wide range of data. Normally it may produce a trade directory and, perhaps, a yearbook.

3. Syndicated reports – These are published market research reports from various organizations which charge for their information. These data relate to consumer information and media information. These are generally prepared to cater to all interested and not to any specific organizational requirement. Hence the relevant data is extracted from this.

4. National and international institutions: Economic reviews, Research reports, journals and articles are all useful sources to contact. A lot of secondary data can be obtained from World Bank, WHO, International Monetary Fund, International Fund for Agricultural Develoonment, United Nations Development Programme, Food and Agriculture Organization and ILO.

4.3 PRIMARY DATA

Primary data collection begins when a researcher is not able to find the data required for his research purpose from the secondary sources. The primary data are that, which are collected afresh and for first time and thus happens to be original in character.

Check your Progress

2. What is a primary data?

4.4.1. Methods of collecting primary data

There are three basic ways of collecting primary data. They are:

Observation method

Survey methods

Case study

There are three basic means of obtaining primary data depending on the nature of the problem and by the availability of time and money:

4.4 OBSERVATION

As the name implies, the researcher observes the situation of interest and records the relevant facts, actions, or behaviors. Observation provides accurate data about what consumers do in certain situations but do not provide details of why it happened. Observation becomes a scientific tool when it is systematically planned and recorded and is subjected to checks and controls on validity and reliability.

4.4.1 Advantages

- The method eliminates subjective bias
- The information obtained under this method relates to what is current happening it is not complicated either by past behaviour or future intentions and attitudes.
- This method is independent of respondent willingness to respondents as such is relatively less demanded of active co-operation on the part of the respondents as happens to be the case in interview or the questionnaire method.
- This method is particularly suitable in studies, which deal with subjects who are not capable giving verbal reports of their feeling for one reason or the other.

4.4.2 Disadvantages

- Its s an expensive method

The information provided by this method is very limited.

- Sometimes unforeseen factors may interfere with the observational task.
- The fact that some people are rarely accessible to direct observation creates obstacle for this method to collect data effectively.

4.4.3 Methods of observation:

There are several methods of observation:

- **Structured** – unstructured observation: In structured observation, the observer is given a set of behaviors to observe. In unstructured observation, the observer is allowed to observe anything that may be relevant to the research objective.
- **Disguised** – undisguised observation: In a disguised observation, the subjects do not know that they are being observed. This is a better way to observe as the subjects are not conscious that they are being observed and behave freely.
- **Observation under normal setting** – Laboratory setting: Normal setting would be a field survey; laboratory setting would be under a fixed roof or venue. The latter allows a prompt and economical way of collecting data and permit the use of more objective measurements.
- **Direct – Indirect Observation** – In the case of direct observation, the event or the behaviour of a person is observed as it occurs. In an indirect observation, a record of a past behaviour is observed.
- **Human – Mechanical Observation** – The observations are recorded manually in Human Observations. In Mechanical Observation, the research is carried out through hidden cameras and audiometers; hence there is no subjective bias.

Check your Progress

3. What is a structured observation?

4.5 SURVEYS

A survey is a research technique in which data are systematically collected directly from the people being studied by the questionnaire. Surveys are a form of questioning that is more rigid than interviews and that involve larger groups of people. Surveys will provide a limited amount of information from a large group of people and are useful when you want to learn what a larger population thinks.

4.5.1 Types of survey

a) Telephone survey

b) Mail survey

c) E-mail survey

d) Internet, Intranet, web page surveys

a. Telephone Surveys

Surveying by telephone is the most popular interviewing method in advanced countries. Where the time is short and distance is too far the research questions will be asked through telephone.

Advantages

1. People can usually be contacted faster over the telephone than with other methods. If the Interviewers are using CATI (computer-assisted telephone interviewing), the results can be available minutes after completing the last interview.

2. You can dial random telephone numbers when you do not have the actual telephone numbers of potential respondents.

Disadvantages

1. Many people are reluctant to answer phone interviews and use their answering machines to screen calls.

2. You cannot show your sample products by phone.

b. Mail Surveys

Mail and telephone surveys are a method of collecting information by sending surveys via email or postal mail. Participants return completed forms to the researcher. Surveys may ask respondents to rate items on a scale .Some surveys also allow respondents to write their feelings or attitudes about a particular event or to elaborate in more detail on an item, or to express suggestions, etc.

Advantages

1. Mail surveys are among the least expensive.
2. This is the only kind of survey you can do if you have the names and addresses of the target population, but not their telephone numbers.
3. The questionnaire can include pictures - something that is not possible over the phone.
4. Mail surveys allow the respondent to answer at their leisure, rather than at the often inconvenient moment they are contacted for a phone or personal interview. For this reason, they are not considered as intrusive as other kinds of interviews.

Disadvantages

1. Time- Mail surveys take longer than other kinds. You will need to wait several weeks after mailing out questionnaires before you can be sure that you have gotten most of the responses.
2. In populations of lower educational and literacy levels, response rates to mail surveys are often too small to be useful.

c. Email Surveys

Email surveys are both very economical and very fast. More people have email than have full Internet access. This makes email a better choice than a Web page survey for some

populations. On the other hand, email surveys are limited to simple questionnaires, whereas Web page surveys can include complex logic.

Advantages

1. Speed. An email questionnaire can gather several thousand responses within a day or two.
2. There is practically no cost involved once the set up has been completed.
3. You can attach pictures and sound files.
4. The novelty element of an email survey often stimulates higher response levels than ordinary "snail" mail surveys.

Disadvantages

1. You must possess (or purchase) a list of email addresses.
2. Some people will respond several times or pass questionnaires along to friends to answer. Many programs have no check to eliminate people responding multiple times to bias the results. The Survey System's Email Module will only accept one reply from each address sent the questionnaire.
3. Many people dislike unsolicited email even more than unsolicited regular mail. You may want to send email questionnaires only to people who expect to get email from you.
4. You cannot use email surveys to generalize findings to the whole populations. People who have email are different from those who do not, even when matched on demographic characteristics, such as age and gender.
5. Email surveys cannot automatically skip questions or randomize question or answer choice order or use other automatic techniques that can enhance surveys the way Web page surveys can.

d. Internet/Intranet (Web Page) Surveys

Web surveys are rapidly gaining popularity. They have major speed, cost, and flexibility advantages, but also significant sampling limitations. These limitations make software selection especially important and restrict the groups you can study using this technique.

Advantages

1. Web page surveys are extremely fast. A questionnaire posted on a popular Web site can gather several thousand responses within a few hours. Many people who will respond to an email invitation to take a Web survey will do so the first day, and most will do so within a few days.
2. There is practically no cost involved once the set up has been completed. Large samples do not cost more than smaller ones (except for any cost to acquire the sample).
3. You can show pictures. Some Web survey software can also show video and play sound.
4. Web page questionnaires can use complex question skipping logic, randomizations and other features not possible with paper questionnaires or most email surveys. These features can assure better data.
5. Web page questionnaires can use colors, fonts and other formatting options not possible in most email surveys.
6. A significant number of people will give more honest answers to questions about sensitive topics, such as drug use or sex, when giving their answers to a computer, instead of to a person or on paper.
7. On average, people give longer answers to open-ended questions on Web page questionnaires than they do on other kinds of self administered surveys.

Disadvantages

1. Current use of the Internet is far from universal. Internet surveys do not reflect the population as a whole.
2. People can easily quit in the middle of a questionnaire. They are not as likely to complete a long questionnaire on the Web, as they would be if talking with a good interviewer.
3. If your survey pops up on a web page, you often have no control over who replies.
4. Depending on your software, there is often no control over people responding multiple times to bias the results.

4.5.2 Things to consider when conducting surveys:

Who are you planning on surveying? Decide what group you are going to focus on surveying based on who you have access to and what your research is focused on. How many people are you going to survey? You want to choose a target number of surveys to conduct. You don't want too few surveys because you won't have enough answers to support any generalizations or findings you may make. At the same time, you do not want too many surveys because you will be overwhelmed with analyzing your data. How are you going to survey people? You can choose to conduct your survey in person (i.e. walk up to people and ask them questions); on paper (i.e. hand out surveys and ask people to return them); or even via the Internet. The survey method should be chosen based on the length of your survey and types of questions. How long is your survey going to be? The answer to this question depends on what information you are attempting to discover and how much you want to find out. Longer surveys sometimes involve the same question asked in multiple ways to see if people are consistent in their answering strategies. For your first survey, however, it is better to keep things simple. Short

questions are usually more effective than longer ones. What type of questions are you going to ask? Do you want open-ended questions or closed questions? Open-ended questions are questions that allow the participant any type of response. An example of an open-ended question is: How are you feeling today? A closed question is one with a set of possible responses or yes/no responses. An example is:

Did you feel that the new campus regulation about parking is fair? While closed questions are much easier to analyze they do not provide the rich responses you may get with open-ended questions. Ultimately, what type of question you ask depends on what you want to discover.

Comparison of Survey Characteristics

Survey characteristics	**Mail survey**	**Telephone survey**	**Personal interview survey**
Cost	Lowest	Moderate	High
Geographic distribution	Wide	Wide	Moderate
Flexibility in questioning	Low	Moderate	High
Interviewer bias	No	Moderate	High
Speed and data collection	Slowest	Fastest	Moderate
Control and data collection	Lowest	Moderate	Highest
Response rates	Poor	Moderate	Highest

4.6 CASE STUDY

Case study is a very popular method of qualitative research. It involves a careful and complete observation rather than using large samples and following a rigid protocol to examine a limited number of variables. case study methods involve an in-depth. longitudinal examination of a single instance or event: a case. They provide a systematic way of looking at events. collecting data, analyzing information, and reporting the results. As a result the researcher may gain a sharpened understanding of why the instance happened as it did, and what might become important to look at more extensively in future research. Case studies lend themselves especially to generating (rather than testing) hypothesis The design involves the intense investigation of situations which are relevant to the problem situation. The case history method is especially useful in situations in which a complicate series of variables interact to produce the problem or opportunity. The concept is to select several target cases where an intensive analysis will help to

- Know the environment
- Identify relevant variables
- Indicate nature of relationship among the variable
- Identify the nature of problems and opportunities in the original case.

Check your Progress

4. What is a case study?

4.7 PERSONAL INTERVIEW

An interviewer asking questions generally face-to-face to other persons conducts personal interview. This sort of interview may be in the form of direct personal investigation or it may be an indirect oral investigation. This method is particularly suitable for intensive investigations.

Advantages

- More information and that too in greater depth can be obtained.
- Interviewer can overcome any resistance, if any, of the respondents; this interview can be made to yield an almost perfect sample of the population.
- There is greater flexibility as questions can be restructured as when needed, especially in the unstructured interviews.
- Observation method can supplement verbal recording of answers.
- Personal information can be obtained easily in this method.
- Sample control can be maintained, as non-response generally remains low.
- Unlike mailed questionnaire, the interviewer can usually control which persons will answer the questions.
- The interviewer can catch the respondent off-guard and thus record the spontaneous reactions.
- The language of the interview can be changed according to the education level of the respondent.
- The interviewer can collect supplementary information about respondent's personal characteristics and environment, which helps while interpreting, results.

Disadvantages

- It can be quite expensive method, especially when large and widespread geographical sample is taken.
- Possibility of bias of interviewer and respondent is maximum.
- Certain respondents such as important officials cannot be approachable under this method.

Check your Progress

5. What are the differences between a questionnaire and a schedule

It is time-consuming especially when sample is large and re-calls o respondents are to be made.

Sometimes the presence of he interview can over-stimulate he respondent and he may give imaginary answers to make the interview interesting.

Under the interview method the organization required for selection, training and supervising the field-staff is more complex with formidable problems.

Interviewing at times may introduce systematic errors.

Interview presupposes a proper rapport with respondents for free and frank responses, which is not always possible.

4.7.1. Telephone interviews

This method of collecting information consists contacting information consists contacting respondents on telephone itself. It is not a very widely used method, but plays important role in industrial surveys in developed regions.

Advantages

It is more flexible in comparison to mail method.

It is faster in obtaining information than other methods.

It is cheaper compared to personal interviews; here the cost per response is very low.

Recall is easy; callbacks are economic and simple.

There is higher rate of response than mailing method

Replies can be recorded without causing embarrassment to respondents.

Interviewer can explain requirements more easily.

Access can be gained to respondents who otherwise cannot be contacted for one reason to other.

No field staff is required.

Wider distribution of sample is possible.

Disadvantages

➢ Little time is given to respondents to answer, as these types of interviews do not last for more than 5 minutes.

➢ Survey is restricted to people who have telephones.

➢ Cost plays a major part in extensive geographical coverage.

➢ It is not suitable for interviews having comprehensive answers to various questions.

➢ Some extent of interviewer's bias exists.

➢ Questions have to be short and probes are difficult to handle.

4.8 DATA COLLECTION INSTRUMENTS

Data collection instruments used by survey method are:

a) Questionnaires

b) Commercial surveys

c) Audit

d) Panels

e) Mail questionnaires

a. Questionnaire

Questionnaires are data collecting instruments used to ask respondents questions to secure the desired information. Questionnaires may be administered by mail, over the telephone, by computer, or in person. The design of a questionnaire depends on whether the researcher wishes to collect exploratory information (i.e. qualitative information for the purposes of better understanding or the generation of hypotheses on a subject) or quantitative information (to test specific hypotheses that have previously been generated).The questionnaires can be classified into four types:

1. Structured – non-disguised
2. Structured disguised
3. Non-structured - non- disguised

4. Non-structured – disguised

Non-disguised are the direct questions and the object of enquiry is revealed to the respondent. Disguised are the indirect questions where the object of enquiry is not revealed to the respondent. In a structured questionnaire the questions are asked in a pre-determined order. Formal standardised questionnaires: If the data is required to be analysed statistically, a formal standardised questionnaire is designed.

Development of a questionnaire:

The following steps are involved in the development of a questionnaire:

1. Choose and decide on the information required
2. Define the target respondents
3. Selection of methodology to reach the target segment
4. Decide on question content

Thus while forming the question content the following question must be answered appropriately.

Is the Question Necessary/Useful?

Do Respondents Have the Needed Information?

Does the Question Need to be More Specific?

Is Question Biased or Loaded?

Will Respondents Answer Truthfully?

5. Decide on type of questions

The way questions are phrased is important and there are some general rules for constructing good questions in a questionnaire.

- Use short and simple sentences
- Short, simple sentences are generally less confusing and ambiguous than long, complex ones. As a rule of thumb, most sentences should contain one or two clauses.
- Ask for only one piece of information at a time
- Avoid negatives if possible

Ask precise questions

Ask for the exact level of details required.

Minimize bias

6. Putting questions into a meaningful order and format

In order to make the questionnaire effective and to ensure quality to the replies received, a researcher must pay attention to the question-sequence in preparing the questionnaire. A proper question sequence reduces the chances of the questions being misunderstood. The order of the questions is also important. Some general rules are:

-Go from general to particular.

-Go from easy to difficult.

-Go from factual to abstract.

-Start with closed format questions.

-Start with questions relevant to the main subject.

-Do not start with demographic and personal questions.

Opening questions: Opening questions generally should be easy to answer and not in any way threatening to the respondents. This is crucial because it is the respondent's first exposure to the interview and sets the tone for the nature of the task to be performed. If they find the first question difficult to understand, or beyond their knowledge and experience, or embarrassing in some way, or uninteresting they are likely to break off immediately. If, on the other hand, they find the opening question easy and pleasant to answer, they are encouraged to continue.

Question flow: Questions should flow in some kind of psychological order, so that one leads easily and naturally to the next. There could be a continuity maintained on the flow of the questions where the response from on leads to another. This helps in creating a sequence and the respondent's interest is

maintained. Questions on one subject, or one particular aspect of a subject, should be grouped together. Respondents may feel it disconcerting to keep shifting from one topic to another, or to be asked to return to some subject they thought they gave their opinions about earlier.

Question variety: Respondents become bored quickly and restless when asked similar questions for half an hour or so. Hence the questions need to be carefully keyed in to maintain the interest throughout the interview.

7. Closing questions

By the time the respondent comes to the end of the questionnaire it is quite natural for a respondent to become increasingly indifferent to the questionnaire. This is mainly due to impatience or fatigue. He might give careless answers to the later questions. Hence such questions should be included in the earlier part of the questionnaire. Potentially sensitive questions should be left to the end, to avoid respondents cutting off the interview before important information is collected.

8. Physical appearance of the questionnaire

Data quality can also be affected by the physical appearance of the questionnaire with unnecessarily confusing layouts making it more difficult for interviewers, or respondents in the case of self-completion questionnaires, to complete this task accurately. In general it is best for a questionnaire to be as short as possible. A long questionnaire leads to a long interview and this may lead to decreasing interest on the part of the respondent.

Piloting the questionnaires

Piloting is very much essential to test whether the desired responses are being obtained for the questions. Many a time, the perception of the respondents varies from those of the

researcher. Hence these issues can be corrected in the initial stage itself so that the research process is facilitated. The purpose of pretesting the questionnaire is to determine:

- If the wordings used help in achieving the desired results
- Are the questions in the right order?
- Are the questions easy to understand?
- If any questions needed to be added or deleted.
- Are the instructions to interviewers are adequate?

b. Commercial Surveys

Commercial surveys can be divided into three types: Periodic, Panel and Shared surveys. Each of them are discussed below

Periodic surveys

Periodic surveys are conducted at regular intervals, ranging from weekly to annually held surveys. They use a new sample of respondents for each survey, focusing on the same topic and allowing the analysis of trends over a period. Periodic surveys are conducted by mail, personal interview and telephone. The disadvantage here could be that when periodic surveys are conducted at known intervals, they might affect the behavior being measured. An example of this kind of surveys could be TRPs.

Panel surveys

Panel surveys, sometimes called interval panels, are conducted among a group of respondents who have agreed to respond to a number of mail, telephone or occasionally personal interviews over time. These need not occur regularly. But a continuous panel or panel data (explained more in panels) refers to a group of individuals who agree to report specified behaviors over time. The advantages of this method are:

The research firm initially collects all the personal information about the respondents and does not waste time again in collecting this information during interviews. This increases the quality of the research data. The response rate can be as high as 70% - 90%.

Shared surveys

Shared surveys, sometimes referred to as omnibus surveys, are administered by a research firm and consist of questions supplied by multiple clients. Such surveys can involve mail, telephone, or personal interviews. The respondents may be drawn from either an interval panel or random selection.The main advantage here is the cost factor.

c. Audits

Audits involve the physical inspection of inventories, sales receipts, shelf facing and other aspects of marketing mix to determine sales, market share, relative price, distribution and other relevant information. The different types of audits are store audits, product audits and retail distribution audits.

Store audits

The basis for the store audit of retail stores sales is the simple accounting arithmetic of = [(Opening inventory + Net purchases (receipts-transfers out-returned inventory + transfers in) - Closing inventory]/ Sales

These audits provide sales data on packaged products. The clients receive report on the sales of their own brand and of competitor's brands, the resulting market shares, prices, shelf facing and other important information.

Product audits

Product audits are similar to store audits but focuses on products rather than store samples. Although they provide similar information as of store audits it differs as in it tries to

cover all the types of retail outlets that handle a product category.

Retail distribution audits

Retail distribution audits are similar to store audits however these audits do not measure inventory sales: instead they are observational studies at the retail level. Field agents enter stores unannounced and without permission. They observe and record the brands present, price, shelf facings and other relevant data for selected product categories.

d. Panels

A panel is a group of individuals or organizations that have agreed to provide information to researcher over a period of time. A continuous panel, the focus of this section, has agreed to report specified behaviors on regular basis. There are 2 types of panels: retail and consumer, consumer further divided into diary panels and electronic panels.

Retail panel

In this method data is collected from the checkout scanner tapes of a sample of supermarkets and other retailers that use electronic scanning systems. For this to happen the product should carry the Universal Product Code (UPC) often referred to as bar code. The advantages of this method are

- Greater frequency
- Elimination of breakage and pilferage being counted as sales
- More accurate price information
- The disadvantages are
- Only big supermarkets have scanners

The quality of scanner data is dependent on checkout clerk. For e.g. if a person is buying 5 packets of packaged milk. In that

case the clerk may put only 1 in the scanner and then multiply it by 5. So the rest 4 wont come in the scanner's data.

Consumer panels

a. Diary panels

A diary panel as the name implies, is a panel of households who continuously record in a diary their purchase of selected products. It is used for those product categories for which purchasing is frequent like food and personal care products.

b. Electronic panels

Electronic panels are composed of households whose television viewing behavior is recorded electronically. The sets were wired to household meters. The meters were connected to a central computer by a telephone line and automatically recorded when the set was turned on and the station to which it is turned on. The problem here is that it is difficult to understand whom all and how many people were watching and what their demographics are.

e. Mail Questionnaire

Advantages

- It is easier to approach a large no. Of respondents spread all over the world through post.
- A mail questionnaire is free from any interviewer's bias and errors, which may undermine the reliability and validity of the results emerging from the survey.
- A mail questionnaire will not have any distribution bias as it will not show any particular preference or dislike for a certain individual or household.
- When the questions asked to the respondents need time to be answered and needs some thinking, mail questionnaire is ideal.

- Mail Questionnaire saves time in collecting the desired information as a large no. Of respondents can be approached all over the country.
- It saves money as cost of traveling, boarding and lodging of interviewers is not to be incurred.
- There is no difficulty in having central supervision and control over the survey operations over a large region.
- It avoids the bias arising from any inhibitions in answering questions. (During some personal questions the respondents may hesitate to answer them in the presence of the interviewer)
- It will not have the problem of non-contacts in the strict sense, as might be the case in personal interviews when the interviewer finds that the respondent, being away from home is not available.

Limitations

- It is not suitable when questions are difficult & complicated. In such a case the help of interviewer is required to offer some introductory explanation to the respondent.
- When the objective is to get the spontaneous answers of the respondent or his own answers uninfluenced by others who may influence his thinking.
- It is not possible to verify whether the respondent himself has filled in the questionnaire. (e.g.: If a questionnaire is targeted to a housewife she may ask her husband to fill it up on her behalf). This can result into incorrect answers.
- In case there is any ambiguity or any inconsistency in the answers it will be difficult for the researcher to make use of such questionnaire, as he has to accept it.
- The respondent may go through his answers after he has filled in the entire questionnaire and may make certain

modification in his original answers as a result of which these answers cannot be regarded as independent.

- It does not allow the researcher to supplement the information by his personal observations.
- A mail questionnaire normally has a relatively poor response compared to a questionnaire canvassed personally.

Other Primary data collection Methods

a) Warranty cards

b) Distributors audits

c) Pantry audits

d) Consumer panels

e) Using mechanical devices

f) Through projective techniques

g) Depth interviews

h) Content analysis

4.9 ERRORS AND DIFFICULTIES IN DATA COLLECTION

Since the usability of any market research depends upon the accuracy of the results, error control plays a critical role in the research process. The common marketing research errors are highlighted below. The major types of errors are:

a) Sampling errors

b) Non sampling errors

4.9.1 Sampling errors

Most marketing research studies utilize samples of people, product or stores. Based upon these sample results, the researcher and the manager make conclusion of the whole population from which the sample was selected. For example the attitudes of all Maruti car owners could be inferred from a sample of a 1000 owners because the sample is used to estimate

the population. The difference between the sample value and the corresponding population value is called as sampling error .

4.9.2 Non sampling errors

Non sampling errors are all the errors that may occur in the marketing research process except the sampling error. This includes all the aspects of the research process where mistakes and deliberate deceptions can occur. Unfortunately these mistakes and deceptions occur with great frequency in the marketing research process.

4.9.3 Types of Non Sampling Errors:

- Defective problem definition: Problem of the study should be clearly stated so that they can be linked directly to the research results. Research objectives should always be clear so that research results can be presented in relation to specific objectives.
- Defective population definition: The study population must be defined to fit the study objectives. A universe which is relevant to the problem being studied, and a sample which adequately represents that universe, are vital requirements of high quality research.
- Frame non representative of the population: The sampling frame must match the defined population. Consider the case of the investment company that uses the telephone book (the frame)to select a sample of "potential stock buyers" this frame would not cover the defined population well as a significant number of high income people have unlisted phone numbers.
- Non responsive errors: Errors occur because people in selected sample either refuse to be a part of the sample or they are not at home during the sample periods. The resulting error is called non–response error.

➢ Questionnaire structure error: The error made when the structure and layout of the survey instrument leads to inaccurate responses.

➢ Measurement error: This is caused when information gathered is different from the information sought.

➢ Data analysis error: The error that occurs when analysis is incorrectly executed. Simple mathematical errors are common, which is why data analysis should be checked over by more than one qualified person for quality.

➢ Reporting error: The best approach and program design combined with the best analysis is only as good as the researcher's capability to synthesize and report on the results. The most common reporting error by far is the improper representation of the significant findings in a format conducive to creating management understanding and buy-in of survey results.

4.9.4 Minimizing of Errors

In research it is not possible to eliminate all the errors. However, to the extent that can be minimized. The following precautions can be followed to reduce the errors:

- ➢ Selection of suitable study method
- ➢ Selection of appropriate instrument
- ➢ Adequate sample size
- ➢ Using of trained and experiences researchers
- ➢ Planned data processing

4.10 DIFFICULTIES IN DATA COLLECTION

Marketing research basically is a problem solving tool. Starting from the definition of the problem till the end of the presentation of the report researchers are facing numerous problems. The major problems faced by the researchers are mentioned below:

1. Volatile changes in the business environment make marketing research more complicated.

2. Lack of scientific training.

3. Decisions makers do not appear to be very eager on implement the findings of the study.

4. Many of the organizations feel that investment in research is wastage of resources and does not encourage research.

5. Many people largely depend on customs, traditions and routine practices in their decision making, as they feel research does not have any useful purpose to serve in the management of their business.

6. Most of the business organizations in our country do not part with information to researchers.

7. There is insufficient support between the business organizations and research institutions.

8. In India companies are not in a position to allocate huge funds for the research.

9. Lack of availability secondary data makes marketing research as baseless start and time consuming one.

10. Poor awareness among the consumers about research makes the study more burning.

11. Researchers in India are not familiar with the new research instruments available for conducting market research,

12. Poor library facilities mean researchers have to spend much of time and energy in tracing out relevant material and information.

13. There is a difficulty of timely availability of up to date data from published sources

14. Lack of code of conduct among the researchers brings bad image on research. There is a need for developing code of conduct for researchers.

4.11 SUMMARY

One of the most important components of Marketing Research is collection of data required to solve a defined research problem. The data collection tools help to achieve them

4.12 ANSWERS TO 'CHECK YOUR PROGRESS'

1. A data collected by somebody for some other purpose is secondary data.
2. A first hand data free from boas and editing is primary data.
3. A structured observation has defined structured items to be observed.
4. A case study is a comprehensive study of a social unit , the unit may be a person , or a society, or a company.
5. Research schedule are being filled by the enumerators who are specially appointed for this purpose. Research questionnaire is the set of questions which is mailed to the respondent, respondent will send the filled questionnaire.

4.13 EXERCISES AND QUESTIONS

1. Tabulate the difference between Primary and Secondary data.
2. Does secondary data have some disadvantages? Explain.
3. What is primary data? How is it collected?
4. Explain the methods of observation used to collect primary data.
5. Describe the various steps involved in the development of a questionnaire.

4.14 FURTHER READINGS

Green, Paul E, and Donald S. Tull, Research For Marketing Decisions, Englewood clifs N.J, Prentice Hall, Inc, 2006

G.C.Beri, Marketing Research, TMH, 2008.

Nargundkar R, Marketing Research Text and Cases, TMH, 2002

NOTES

UNIT V

SAMPLING DESIGN

STRUCTURE

5.1 Introduction
5.2 Sampling Plan
5.3 Sampling methods
5.4 Probability sampling methods
5.5 Non-probability sampling methods
5.6 Sample size determination
5.7 Summary
5.8 Answers to 'Check Your Progress'
5.9 Exercises and Questions
5.10 Further Readings

UNIT OBJECTIVES

In this section, you should be able to understand:

- The concept of sampling.
- How an appropriate sampling design can be determined.
- Sampling plan and the steps in developing it.
- Various probability and non-probability sampling methods.
- The concept of sample size and the methods of determining it.

5.1 INTRODUCTION

Sampling involves selecting a relatively small number of elements from a larger defined group of elements and expecting that the information gathered from the small group will allow judgments to be made about the larger group. If all the respondents in a population are asked to provide information, such survey is called a census. Information obtained from a subset of the population is known as the statistic (from sample). Researchers then attempt to make an inference about the population parameter with the knowledge of the relevant sample statistic.

NOTES

5.2 SAMPLING PLAN

A sampling plan is the blueprint or frame work needed to ensure that the raw data collected are representative of the defined target population. A good sampling plan will include the following steps:

Step 1: Define the target population: In any sampling pan, the first task of the researcher is to determine and identify the complete group of people or objects that should be investigated in the project. The target population should be given its identity by the use of descriptors that represent the characteristics of elements that make the target population's frame. These elements become the prosperity sampling units from which a sample will be drawn. Clear understanding of the target population will help the researcher successfully draw a representative sample. Devoting effort to identifying the target population usually will pay off.

Step 2: Select the data collection method: Using the information problem definition, the data requirements, and the established research objectives, the researcher must choose a method for collecting the required raw data from the target population elements. Choices include interviewing approach or a self-administered survey. The method of data collection guides the researcher in identifying and securing the necessary sampling frame(s) for conducting the research.

Step 3: Identifying the sampling frame(s) needed: After gaining an understanding of whom or what should be investigated, the researcher must assemble a list of eligible sampling units. This list needs to contain enough information about each prospective sampling unit so that the researcher can successfully contact them. An incomplete sampling frame decreases the likelihood of drawing a representative sample. Sampling frame lists can be created from a number of different sources. The researcher must be aware of possible conditions of over-registration and under-registration of the prospective sampling units. These conditions will create sampling gaps or sampling frame errors that decrease the likelihood of being able to draw a representative sample.

Step 4: Select the appropriate sampling method: The researcher must choose between two types of sampling orientations: probability and non-probability. Using a probability sampling method will always yield better and more accurate information about the target population's parameters than will any of the available non-probability sampling methods. Probability sampling has several advantages over non-probability sampling. First, it permits the researcher to demonstrate the sample's representativeness. Second, it allows an explicit statement as to how much variation is introduced, because a sample is used instead of a census of the populations. Finally, it makes possible the more explicit identification of possible biases.

Step 5: Determine necessary sample sizes and overall contact rates: In this step of a sampling plan, the researcher must consider how precise the sample estimates must be and how much time and money are available to collect the required raw data. To determine the appropriate sample size, decisions have to be made concerning (1) the variability of the population characteristic under investigation, (2) the level of confidence desired in the estimates, and (3) the degree of precision desired in estimating the population characteristic. The researcher must decide how many completed surveys will need to enter the data analysis activities of the overall research project.

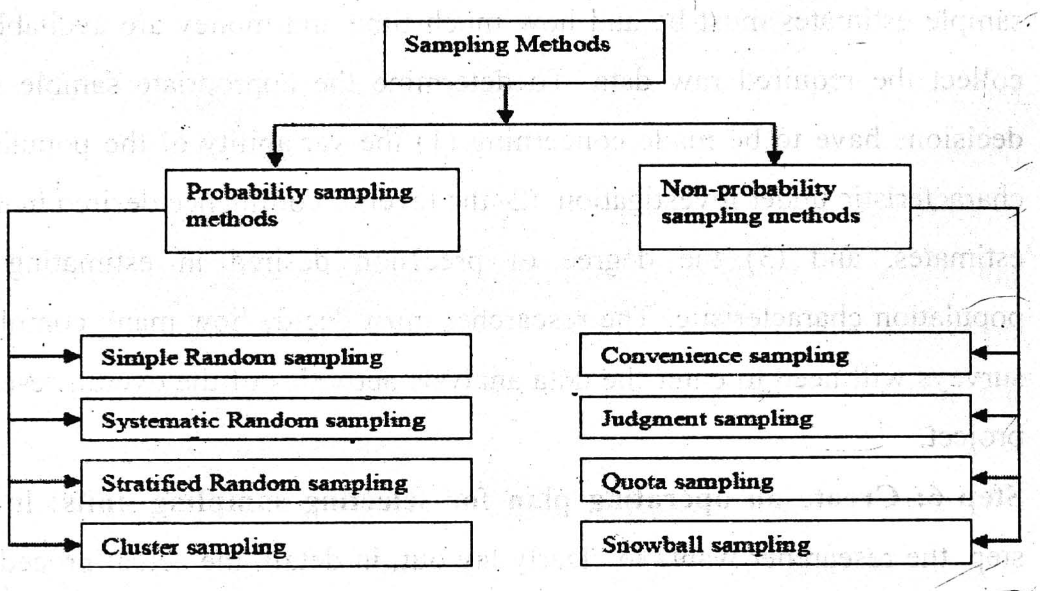
Step 6: Create an operating plan for selecting sampling units: In this step, the researcher wants to clearly lay out, in detail, the actual procedures to use in containing each of the prospective respondents who were drawn into the sample. All instructions should be clearly written so that interviewers know exactly what to do and how to handle any problems in the process of contacting prospective respondents.

Step 7: Execute the operational plan: In some research projects, this step is similar to actually conducting the data collection activities. (e.g., actual calling of a prospective respondent to do a telephone interview). The important thing in this stage is to maintain consistency and control.

5.3 SAMPLING METHODS

There are two basic sampling designs: Probability and non-probability sampling methods. In probability sampling, each unit in the defined target population has a known; non-zero probability of being selected for the sample, the actual probability of selection for each sampling unit may or may not be equal depending on the type of probability sampling design used. It allows the researcher to judge the reliability and validity of raw data collected by calculating the probability to which the findings based on the sample would differ from the defined target population. The results obtained by the probability method can be generalized to the target population within a specified margin of error through the use of statistical methods.

Types of sampling methods



In non-probability sampling, the probability of selection of each sample unit is not known. Therefore, potential sampling error cannot be accurately known either. The selection of sampling units is based on some type of intuitive judgments, desire or knowledge of the researcher. The degree to which the sample may or may not be representative of the defined target population depends on the sampling approach and how well the researcher executes and controls the selection activities. There is always a temptation to generalize non-probability sample data results to the defined target population.

Comparative differences of probability and non-probability sampling methods

Factor sampling	Probability sampling	Non-probability
1.List of the population elements	Complete list necessary	Not necessary
2. Information about the-sampling units	Each unit identified	need detail on habits, activities, traits etc.
3. Sampling skill	skill required	little skill required
4. Time requirement	More time-consuming	Less time consuming
5. Cost per unit sampled	Moderate to high	Low
6. Estimate of population - parameters	Unbiased	Biased
7. Sample representativeness	Assured	Undeterminable
8. Accuracy and Relaiability	Computed with confidence intervals	Unknown
9. Measurement of sampling error	Statistical measures	No true measure available

5.4 TYPES OF PROBABILITY SAMPLING DESIGNS

5.4.1. Simple Random Sampling

Simple Random Sampling is a probability sampling procedure which ensures that every sampling unit making up the defined target population has a known, equal, non-zero chance of being selected. For example, let's say an instructor decided to draw a sample of 10 students (n=10), from among all the students in a Marketing Research class that consisted of 30 students (N=30). The instructor could write each student's name on a separate, identical piece of paper and place all of the names in a jar. Each student would have an equal, known probability of selection for a sample of a given size that could be expressed by the formula:

$$\text{Probability of selection} = \frac{\text{Size of sample}}{\text{Size of population}}$$

Here, each student would have a 10 /30 (or 0.33) chance of being randomly selected in the drawn sample. When the defined target population consists of

a larger number of sampling units, a more sophisticated method would be used to randomly draw the necessary sample. One of the procedures commonly used in marketing research is to incorporate a printed or computer generated table of random numbers to select the sampling units. A table of random numbers is a table that lists randomly generated numbers. Many computer programs have the ability to generate a table random numbers. With the marketing research students defined above as the target population a random sample could be generated by assigning each students a unique two-digit code ranging from 01 to 30. Then we could go to the table of random numbers and select a starting point, which can be anywhere on the table. Using the partial table of random numbers given below, say we select the upper left-hand corner of the table below as our starting point. We would then begin to read down the first column (or across the first row) and select those two-digit numbers that matched the numbers within the acceptable range until 10 students had been selected. Reading down the first column, we would start with 31, and then go to 14, 49, 99, and 54 and so on.

A partial table of random numbers

31 25	81 44	54 34	67 03
14 96	99 80	14 54	30 74
49 05	49 56	35 51	68 36
99 67	57 65	14 46	92 88
54 14	95 34	93 18	78 27
57 50	34 89	99 14	57 37
98 67	78 25	06 90	39 90
40 99	00 87	90 42	88 18
20 82	09 18	84 91	64 80
78 84	39 91	16 08	14 89

Source : M.G.Kendall and B. Babington Smith, "Table of Random Sampling Numbers", Tracts for Computers, 24 (Cambridge, England : Cambridge University Press, 1946), p.33.

Use only those random numbers that matched the numbers within the acceptable range of 01 to 30. Numbers that fall outside the acceptable range would be disregarded. Thus, we would select students with numbers 14, 20, 25, 05, 09, 18, 06, 16, 08, and 30. If the overall research objectives call for

telephone interviews, drawing the necessary sample can be achieved using the random-digit dialing (RDD) technique.

Advantages and disadvantages

The simple random sampling technique has several advantages. The technique is easily understood and the survey's data results can be generalised to the defined target population with a pre-specified margin of error '*e*'. Another advantage is that simple random samples allow the researcher to gain unbiased estimates of the population's characteristics. This method basically guarantees that every sampling unit of the population has a known and equal chance of being selected, no matter the actual size of the sample, resulting in a valid representation of the Defined target population. The disadvantage of this method is the difficulty of obtaining a complete, current, and accurate listing of the population elements. Simple random sampling requires that all sampling units be identified. For this reason, simple random sampling often works best for small populations or those where computer-derived lists are available.

5.4.2. Systematic Random Sampling

Systematic random sampling (SYMRS) is similar to simple random sampling but requires that the defined target population be ordered in some way, usually in the form of a customer list, taxpayer roll, or membership roster. In research practices, SYMRS has become a very popular alternative probability method of drawing samples. Compared to simple random sampling, systematic random sampling is potentially less costly because it can be done relatively quickly. When executed properly, SYMRS can create a sample of objects or prospective respondents that is very similar in quality to a sample drawn using simple random sampling. To employ SYMRS, the researcher must be able to secure a complete listing of the potential sampling units that make up the defined target population. Individual sampling units are selected according to their position using a skip interval. The skip interval is determined by dividing

the number of potential sampling units in the defined target population by the number of units desired in the sample. The required skip interval is calculated using the formula:

$$\text{Skip interval} = \frac{\text{Defined target population list size}}{\text{Desired sample size}}$$

For instance, if the researcher wants a sample of 100 to be drawn from a defined target population of 1,000, the skip interval would be 10 (ie. 1,000/100). Once the skip interval is determined, the researcher would then randomly select a starting point and take every 10th unit until he proceeded through the entire target population list. There are two important considerations when using systematic random sampling. First, it is important that the natural order of the defined target population.

Advantages and disadvantages

Systematic sampling is frequently used because, if done correctly, it is a relatively easy way to draw a sample while ensuring randomness. The availability of lists and the shorter time required to draw a sample makes systematic sampling an attractive, economical methods for researchers. The greatest weakness of systematic random sampling is the potential for there to be hidden patterns in the data that are not found by the researcher. This could result in a sample that is not truly representative of the defined target population. Another difficulty is that the researcher must know exactly how many sampling units make up the defined target population. In research situations in which the size of the target population is extremely large or unknown, identifying the true number of units is difficult, and even estimates may not be accurate.

5.4.3. Stratified Random Sampling

Stratified random sampling (STRS) requires the separation of the defined target population into different groups, called strata, and the selecting of samples from each stratum. The goal in stratifying is to minimize the variability (or skewness) within each stratum and maximize the differences between strata. In some ways, STRS can be compared to

NOTES

segmentation of the defined target population into smaller, more homogeneous sets of elements. To ensure that the sample maintains the required precision of the total population, representative samples must be drawn from each of the smaller population groups. Drawing a stratified random sample involves three basic steps:

i. Dividing the target population into homogeneous sub-groups or strata.

ii. Drawing random samples from each stratum.

iii. Combining the samples from each stratum into a single sample of the target population.

There are two common methods for deriving samples from the strata:

a. Proportionate and

b. Disproportionate

a. Proportional Stratified Sampling

In proportionate stratified sampling, the sample size from each stratum is depended on the stratum's size relative to the defined target population. Therefore, the larger strata are sampled more heavily using proportionate stratified sampling because they make up a larger percentage of the target population. In this type of sampling procedure the number of objects or sampling units chosen from each group is proportional to the number in the population. Proportional stratified sampling can further be classified as

- *Directly proportional* and
- *Inversely proportional stratified sampling.*

Directly Proportional Stratified Sampling

Assume that a researcher is evaluating customer satisfaction for a beverage that is consumed by a total of 600 people. Among the 600 people, 400 are brand-loyal and 200 are variety-seeking. Past research indicates that the level of customer satisfaction is related to consumer characteristics, such as being either brand-loyal or variety-seeking.

Therefore, it should be beneficial to divide the total population of 600 consumers into two groups

400 and 200 each and randomly sample from within each of the two groups. If a sample size of 60 is desired, then a 10 percent directly proportional stratified sampling is employed.

Consumer Type	Group Size	10 Percent Directly Proportional Stratified Sample Size
Brand-loyal	400	40
Variety-seeking	200	20
Total	600	60

Inversely Proportional Stratified Sampling

Assume that among the 600 consumers in the population, say 200 are very heavy drinkers and 400 are light drinkers. If a researcher values the opinion of the heavy drinkers more than that of the light drinkers, more people will have to be sampled from the heavy drinkers group. In such instances, one can use an inversely proportional stratified sampling. If a sample size of 60 is desired, a 10 percent inversely proportional stratified sampling is employed.

Consumer Type	Group Size	10 Percent Inversely Proportional Stratified Sample Size
Heavy Drinkers	400	40
Light Drinkers	200	20
Total	600	60

b. Disproportional Stratified Sampling

In disproportionate stratified sampling, the sample size selected from each stratum is independent of that stratum's proportion of the total defined target population. This approach is used when stratification of the target population produces sample sizes for sub-groups that contradict their relative importance to the study. An alternative type of disproportionate stratified method is optimal allocation. In this method, consideration is given to the relative size of the stratum as well as variability within the stratum to determine the necessary sample size of each stratum. The basic

logic underlying optimal allocation is that the greater the homogeneity of the prospective sampling units within a particular stratum, the fewer the units that would have to be selected to the estimate the true population parameter (*0* or *P*) accurately for that sub- group. In stratified sampling, when the sample size in each group is not proportional to the respective group sizes, it is known as disproportional stratified sampling. When multiple groups are compared and their respective group sizes are small, a proportional stratified sampling will not yield a sample size large enough for meaningful comparisons, and disproportional stratified sampling is used. One way of selecting sample sizes within each group is to have equal group sizes in the sample. In the example of heavy and light drinkers, a researcher could select 30 people from each of the two groups. In general, stratified sampling is employed in many research projects, because it is easy to understand and execute.

Advantages and Disadvantages

Dividing the defined target population into homogeneous strata provides several advantages, including:

i. the assurance of representative ness in the sample;

ii. the opportunity to study each stratum and make relative comparison between strata; and

iii. the ability to take estimates for the target population with the expectation of greater precision or less error in the overall sample.

The primary difficulty encountered with stratified sampling is determining the basis for stratifying. It is imperative that the basis for stratifying be directly associated with the target population's characteristics of interest. Normally, the larger the number of relevant strata, the more precise the results. The inclusion of excess or irrelevant strata will only waste time and money without providing meaningful results.

5.4.4. Cluster Sampling

While cluster sampling is similar to stratified random sampling, it is different in that the sampling units are divided into mutually exclusive and

collectively exhaustive sub-populations, called clusters. Each cluster is assumed to be representative of the heterogeneity of the target population Examples of possible divisions for cluster sampling include the customers who patronize a store on a given day, the audience for a movie shown at a particular time (e.g., the matinee), or the invoices processed during a specific week. Once the cluster has been identified, the prospective sampling units are drawn into the sample by either using a simple random sampling method or canvassing all the elements within the defined cluster.

A popular form of cluster sampling is **area sampling.** In area sampling, the clusters are formed by geographic designations. Examples include cities, sub divisions and blocks. Any geographical unit with identifiable boundaries can be used. When using area sampling, the researcher has to additional options : the one-step approach or the two -step approach. When deciding on using one-step approach, the researcher must have enough prior information about the various geographic clusters. By assuming that all the clusters are identical, the researcher can focus his attention on surveying the sampling units within one designated cluster group and then generalize the data results to the full target population. The probability aspect of this particular sampling method is executed by randomly selecting one geographic cluster and performing a census on all the sampling units located within that selected cluster. Alternatively, the researcher may execute a two-step cluster sampling approach. First, the researcher would randomly sample a set of cluster and then would decide on the most appropriate probability method to sample individuals within each of the selected clusters. The two-step approach is preferable over the one-step approach, because there is a strong possibility that a single cluster will not be as representative of all other clusters as the researcher thinks.

Advantages and Disadvantages

The cluster sampling method is widely used in marketing research due to its overall cost-effectiveness and feasibility of implementation, especially in area sampling situations. In many cases, the only reliable sampling unit

frame available to researcher is one that describes and lists clusters. These lists of geographic regions, telephone exchanges, or blocks of residential dwellings can normally be easily compiled. Clustering method tends to be a cost-efficient way of sampling and collecting raw data from a defined target population. One primary disadvantage related to cluster sampling is the tendency for clusters to be homogeneous. The more homogeneous the cluster, the less precise the derived sample estimate in representing the defined target population's parameters. The actual object or people within a cluster should be as heterogeneous as those in the target population itself. Another concern with cluster sampling methods is the appropriateness of the designated cluster factor used to identify the sampling units within clusters. A comparison between the stratified sampling process and the cluster sampling process is given in the following table:

Stratified Sampling	**Cluster Sampling**
Homogeneity within group	Homogeneity between groups
Homogeneity between groups	Homogeneity within group
All groups are included	Random selection of groups
Sampling efficiency improved by increasing accuracy at a faster rate than cost	Sampling efficiency improved by decreasing cost at a faster rate than accuracy

5.5 TYPES OF NON-PROBABILITY SAMPLING DESIGNS

Non-probability sampling typically is used in situations such as:

i. The exploratory stages of a research project,

ii. Pre-testing a questionnaire,

iii. Dealing with a homogeneous population,

iv. When a researcher lacks statistical knowledge, and

i. When operational ease is required.

5.5.1. Convenience Sampling

Convenience sampling (or accidental sampling) is a method in which samples are drawn at the convenience of the researcher or

NOTES

interviewer. The assumptions are that the target population is homogeneous and the individuals interviewed are similar to the overall defined target population with regard to the characteristics being studied.

Advantages and Disadvantages

Convenience sampling allows a large number of respondents to be interviewed in a relatively short time. For this reason, it is commonly used in the early stages of research. The use of convenience samples in the development phases of constructs and scale measurements can have a seriously negative impact on the overall reliability and validity of those measures and instruments used to collect raw data. Another major disadvantage of convenience samples is that the raw data and results are not generalized to the defined target population with any measure of precision. It is not possible to measure the representativeness of the sample, because sampling error estimates cannot be accurately determined.

5.5.2. Judgment Sampling

In judgment sampling, (also referred to as purposive sampling), participants are selected according to an experienced individual's belief that they will meet the requirements of the study. Judgmental sampling is associated with a variety of biases. For example, shopping center intercept interviewing can over-sample those who shop frequently, who appear friendly, and who have uncertainty, because the sampling frame is unknown and the sampling procedure is not well specified. There are situations where judgmental sampling is useful and even advisable. First, there are times when probability sampling is either not feasible or expensive. For example, a list of sidewalk vendors might be impossible to obtain, and a judgmental sample might be appropriate in that case. Second, if the sample size is to be very small - say, under 10 – a judgmental sample usually will be more reliable and representative than a probability sample. Suppose one or two cities of medium size are to be used to represent 200 such cities. Then it would be appropriate to pick judgmentally two cities that appeared to be most representative with

Check your Progress

1. List the probability sampling methods

respect to such external criteria as demographics, media habits, and shopping characteristics. Third, sometimes it is useful to obtain a deliberately biased sample. If, for example, a product or service modification is to be evaluated, it might be possible to identify a group that, by its very nature, should be disposed toward the modification.

Advantages and Disadvantages

If the judgment of the researcher or expert is correct, then the sample generated from judgment sampling will be much better than one generated by convenience sampling. But, it is not possible to measure the representativeness of the sample. The raw data and information collected from sampling units generated though the judgment sampling method should be interpreted as nothing more preliminary insights.

5.5.3 Quota Sampling

The quota sampling method involves the selection of prospective participants according to pre-specified quota regarding either demographic characteristics (e.g., age, race, gender, income), specific attitudes (e.g., satisfied/dissatisfied, liking/disliking, great/marginal/no quality), or specific behaviours (e.g., regular/occasional/ rare customer, product user/non user). The underlying purpose of quota sampling is to provide an assurance that prespecified sub-groups of the defined target population are represented on pertinent sampling factors that are determined by the researcher or client. Surveys frequently use quotas that have been determined by the specific nature of the research objectives. In order to meet the quotas, researcher using quota sampling sometimes overlook the problems associated with adhering to the quotas. Assume that an oil company is interested in finding out if women assume responsibility for vehicle maintenance. The company is interested in interviewing women aged below 35 and with age equal to and above 35, as well as working women and nonworking women. Suppose the distribution of the population of women in a city (N=1,000) is as follow

NOTES

Population Characteristics

	<35 years	35 years and Above	*Total*	*Percentage*
Working women	300	200	500	50
Non-working women	200	300	500	50
Total	**500**	**500**	**1,000**	**100**
Percentage	50	50	100	

Assume that the researcher is interested in interviewing 100 women from this city and develops a quota system such that 50 percent of the sample should be working women and 50 percent of the sample should also be under 35 years old. A quota matrix can be developed for a sample size of 100.

Sample Characteristics

	<35 years	35 years and Above	*Total*	*Percentage*
Working women	50	0	50	50
Non-working women	0	50	50	50
Total	**50**	**50**	**100**	**100**
Percentage	50	50		

Advantages and Disadvantages

The greatest advantage of quota sampling is that the sample generated contains specific sub-groups in the proportions desired by researchers. In those research projects that require interviews, the use of quotas ensures that the appropriate sub-groups are identified and included in the survey. The quota sampling method may eliminate or reduce selection bias on the part of the field workers. An inherent limitation of quota sampling is that the success of the study will be dependent on subjective decisions made by the researchers. Also, it is incapable of measuring the true representativeness of the sample or accuracy of the estimate obtained. Hence, attempts to generalise data results beyond those respondents who were sampled and interviewed become very questionable and may misrepresent the defined target population.

Check your Progress

1. List the non probability sampling methods

NOTES

5.5.4. Snowball Sampling

Snowball Sampling involves the practice of identifying and qualifying a set of initial prospective respondents who can, in turn, help the researcher identify additional people to be included in the study. This method of sampling is also called referral sampling, because one respondent refers other potential respondents. Snowball sampling is typically used in research situations where:

i. the defined target population is very small and unique, and

ii. compiling a complete list of sampling units is a nearly impossible task.

The snowball method would yield better results at a much lower cost. Here the researcher would identify and interview one qualified respondent, then solicit his or her help in identifying other people with similar characteristics. The main underlying logic of this method is that rare groups of people tend to form their own unique social circles.

Advantages and Disadvantages

Snowball sampling is a reasonable method of identifying and selecting prospective respondents who are members of small, hard-to-reach, uniquely defined target population. It is most useful in qualitative research practices, like focus group interviews. Reduced sample sizes and costs are primary advantages to this sampling method. Snowball sampling definitely allows bias to enter the overall research study.

5.6 SAMPLE SIZE

Determining the appropriate sample size is not an easy task. The researcher must consider how precise the estimates must be and how much time and money are available to collect the required data, since the data collection is one of the most expensive components of a study. Three factors play an important role in determining appropriate sample sizes. They are:

Check your Progress

3. What is a mult i-stage sampling?

i. *The variability of the population characteristic under consideration:* The greater the variability of the characteristic, the larger the size of the sample necessary.

ii. *The level of confidence desired in the estimate:* The higher the level of confidence desired, the larger the sample size needed.

iii. *The degree of precision desired in estimating the population characteristic:*

The more precise the required sample results, the larger the necessary sample size.

5.6.1 Estimating the sample size by traditional methods

There are four traditional approaches to determine the sample size. They are :

i. *Judgmentally / arbitrarily*: The researcher can simply select a sample size arbitrarily or on the basis of some judgementally based criterion.. There may be instances where the sample size represents all that where available at a particular point of time.

ii. *Analysis considerations*: Analysis considerations may decide the sample size. Sample size may be determined from the minimum cell size needed.

iii. *The budget:* In certain cases, the budget may determine the sample size.

iv. *Applying standard error*: Sample size determination is based on specifying the desired precision in advance and then applying the appropriate standard error formula.

Two major classes of procedures are available for estimating the sample size

1. Confidence – interval approach: This is based on the idea of constructing confidence intervals around sample means or proportions.

2. Hypothesis-testing approach: This makes use of both type I error (rejecting a true null hypothesis) and Type II error (accepting a false null hypothesis).

Check your Progress

4. List the four approaches for sample size determination

5.6.2. Confidence – interval approach to determine the sample size

In this method, a confidence interval is constructed around sample based mean or proportion. The standard error formulae are used for this purpose. This can be explained with an example. Consider a researcher my have taken a sample of 100 consumers and noted that their average per-capita consumption of orange juice was 2.6 litres per week. Pat studies

indicate that the population standard deviation T can be assumed to be 0.3 litre. With this information, we can find a range around the sample mean level of 2.6 litres for which some prespecified probability statement can be made about the process underlying the construction of such confidence intervals. Suppose that we want to set up a 95% confidence interval around the sample mean of 2.6 litres. The standard error of the mean can be computed as:

$$\sigma_x = \frac{\sigma}{n} = \frac{0.3}{100} = 0.03$$

From the table, we can find that the central 95% of the normal distribution lies within C 1.96 Z variates. 95% confidence interval ranges from 2.54 to 2.66 litres. Thus the pre- assigned chance of finding the true population mean to be within 2.54 and 2.66 litres is 95%.

<u>*The case of sample mean*</u>

Following are the steps involved:

1. *Specify the amount of error (E) that can be allowed.* This is the maximum allowable difference between the sample mean and the population mean. E defines the interval within which 0 will lie with some pre-specified level of confidence.
2. *Specify the desired level of confidence.* It can be 95%.
3. *Determine the number of standard errors (Z) associated with the confidence level.*
4. *Estimate the standard deviation of the population.* The standard deviation can be estimated by judgment, by reference to other studies or by the use of a pilot sample.
5. *Calculate the sample size using the formula for the standard error of the mean.*

$$\sigma_x = \frac{E}{Z}$$

6. Neglacting the finite multiplier, we solve for n in the formula

$$\sigma_{\bar{x}} = \frac{E}{Z} = \frac{\sigma}{n}$$

7. In general we can find n from the formula

$$n = \frac{\sigma^2 Z^2}{E^2}$$

The case of sample proportion

The procedure for determining sample size for interval estimates of proportion are:

1. *Specify the amount of error that can be allowed.* Suppose that the desired reliability is such that an allowable interval of $p - A = \pm\ 0.05$ is set, the allowable error E is 0.05.

2. *Specify the desired level of confidence.* Suppose that the level of confidence is 95%.

3. *Determine the number of standard errors (Z) associated with the confidence level.*

4. *Estimate the population proportion.* (A). The population proportion can be estimated by judgment, by reference to other studies or by the use of a pilot sample.

5. *Calculate the sample size using the formula for the standard error of the proportion.*

$$\acute{o}p = \frac{E}{Z}$$

Check your Progress

5. State the central limit theorem

6. Neglacting the finite multiplier, we solve for n in the formula

$$\acute{o}p = \frac{E}{Z} = A\ (1 - A\ ) / n$$

7. We can find n from the formula

$$n = \frac{A\ (1 - A)\ Z^2}{E^2}$$

5.6.2. Hypothesis testing approach

Sample size can also be determined by the hypothesis testing approach. For this, an assumed probability of making Type I error (called alpha risk) and the probability of making Type II error (called beta risk) are needed. These risks are based on the hypotheses:

H0 : the null hypothesis

H1 : the alternate hypothesis

In hypothesis testing, the sample results sometimes lead us to reject H0 when its is true. This is a type I error. On other occasions, the sample findings may lead us to accept H0 when it is false. This is a Type II error.

The case involving means

The steps are:

1. Specify the values for the null (H0) and the alternate (H1) hypotheses to be tested in terms of population means, $\mu 0$ and $\mu 1$ respectively.

2. Specify the allowable probabilities (α and β respectively) of Type I and Type II errors. The Type I error is the error of rejecting a true null hypothesis. The Type II error is made if the alternate hypothesis is rejected when it is true. α and β are the allowable probabilities of making those two types of errors respectively.

3. Determine the number of standard errors associated with each of the error probabilities α and β.

4. Estimate the population standard deviation σ.

5. Calculate the sample size that will meet the α and β error requirements. Since two sampling distributions are involved, a simultaneous solution of two equations is required to determine the sample size and critical value that will satisfy both equations. These equations are:

$$\textbf{Critical value} = \mu_0 + Z_\alpha \left[\frac{\sigma}{\sqrt{n}} \right]$$

$$\textbf{Critical value} = \mu_1 - Z_\beta \left[\frac{\sigma}{\sqrt{n}} \right]$$

6. Setting the right hand side of these two equations equal and solving for n gives

$$n = \frac{(Z_\alpha + Z_\beta)^2 \sigma^2}{(\mu_1 - \mu_0)^2}$$

The case involving proportions

The steps are:

1. Specify the values for the null (H_0) and the alternate (H_1) hypotheses to be tested in terms of population proportions, π_0 and π_1 respectively.
2. Specify the allowable probabilities (α and β respectively) of Type I and Type II errors.
3. Determine the number of standard errors associated with each of the error probabilities Z_α and Z_β
4. Calculate the desired sample size n from the formula:

$$n = \left(\frac{Z_\alpha \sqrt{[\pi_0 (1-\pi_0)]} - Z_\beta \sqrt{[\pi_1 (1-\pi_1)]}}{\pi_1 - \pi_0} \right)^2$$

Bayesian approach to sample size determination

Bayesian procedures are based on the central principle that one should select the sample size that results in the largest positive difference between the expected payoff of sample information and the estimated cost of sampling. The difference between the expected payoff of the sample information and the estimated cost of sampling is frequently referred to as the expected net gain from sampling. An equivalent way of stating the principle is that one should select the sample size that leads to the largest expected net gain from sampling. In a decisional situation in which one of the primary objectives is to maximize payoff, this rule is appropriate. The general approach to applying it requires the decision maker to:

- Determine the expected value of the sample information for a given sample size.

  - Estimate the sampling cost for that specific option.

  - Find the expected net gain from sampling under that option.

- Search through other sample sizes to find the one that leads to the highest expected net gain from sampling.

While logically sound concept, the Bayesian approach is difficult to implement. The primary problem comes in operationalising the first of the steps stated above. In order to determine the expected value of the sample information for a given sample size, one must relate the sample size being considered to the conditional probabilities of making errors, including the effects of non-sampling errors. In real life situations, this may become very difficult to do.

5.7 SUMMARY

Sampling is an important concept that we practice in our every day life. Hope you students would have attained good awareness about the various sampling designs.

5.8 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Various probability sampling methods

   a. Simple random sampling :

   b. Systematic Random Sampling :

   c. Stratified Random Sampling :

   d. Cluster Sampling :

   e. Area sampling :

2. Various Non-probability sampling methods

   a. Convenience Sampling :

   b. Judgment Sampling :

   c. Quota Sampling :

   d. Snowball Sampling :

3. Multistage Sampling:

As the name suggests, multi stage sampling is carried out in different stages. In each stage progressively smaller (population) geographic areas will be randomly selected.

4. List of the four approaches for sample size determination

1. Judgment method
2. Analytical method
3. The budget
4. Standard error method

5. Central limit theorem

The theorem states that for almost all defined target populations, the sampling distribution of the mean or the percentage value derived from a simple random sample will be approximately normally distributed, provided that the sample size is sufficiently large. When '*n*' is greater than or equal to 30, the sample is a large sample.

5.9 EXERCISES AND QUESTIONS

1. Define sampling.
2. Give the differences between census and sampling.
3. What are the factors influencing the selection of a sampling design ?
4. Define sampling plan. What are the steps in developing a sampling plan ?
5. Differentiate between probability and non-probability sampling methods.
6. What are the types of probability sampling methods ?
7. What are the types of non-probability sampling methods ?
8. What are the factors in determining appropriate sample sizes ?
9. What are the traditional approaches in determining the sample size ?

5.10 FURTHER READINGS

1. Paul. E. Green. Donald. S. Tull. Gerald Albaum. *Research for Marketing Decisions*. Prentice Hall of India Pvt. Ltd. New Delhi.

2. David. A. Aaker. V. Kumar. George.S.Day. *Marketing Research*. John Wiley & Sons Inc. Singapore.

UNIT VI

ATTITUDE MEASUREMENT AND SCALING

STRUCTURE

6.1 Introduction
6.2 Measurement and scaling
6.3 Attitude measurement techniques
6.4 Non disguised, non structured techniques
6.5 Disguised, non structured techniques
6.6 Sampling methods
6.7 Non disguised, structured techniques
6.8 Types of scaling techniques
6.9 Constant sum scales
6.10 Thurstone scale
6.11 Disguised, structured techniques
6.12 Summary
6.13 Answers to 'Check Your Progress'
6.14 Exercises and Questions
6.15 Further Reading

UNIT OBJECTIVES

After studying this unit, you should be able to:

- explain the concepts of measurement and scaling,
- discuss four levels of measurement scales,
- classify and discuss different scaling techniques, and
- Select an appropriate attitude measurement scale for your research problem.

6.1 INTRODUCTION

The qualitative information must be converted into numerical form for further analysis. This is possible through measurement and scaling techniques. A common feature of survey based research is to have respondent's feelings, attitudes, opinions, etc. in some measurable form. For

example, a bank manager may be interested in knowing the opinion of the customers about the services provided by the bank. Similarly, a fast food company having a network in a city may be interested in assessing the quality and service provided by them. As a researcher you may be interested in knowing the attitude of the people towards the government announcement of a metro rail in Delhi. In this unit we will discuss the issues related to measurement, different levels of measurement scales, and various types of scaling techniques and also selection of an appropriate scaling technique.

6.2 MEASUREMENT AND SCALING

Before we proceed further it will be worthwhile to understand the following two terms: (a) Measurement, and (b) Scaling.

a) **Measurement:** Measurement is the process of observing and recording the observations that are collected as part of research. The recording of the objects according to certain prescribed rules. The respondent's, characteristics are feelings, attitudes, opinions etc. For example, you may assign '1' for Male and '2' for Female respondents. In response to a question on whether he/she is using the ATM provided by a particular bank branch, the respondent may say 'yes' or 'no'. You may wish to assign the number '1' for the response yes and '2' for the response no. We assign numbers to these characteristics for two reasons. First, the numbers facilitate further statistical analysis of data obtained. Second, numbers facilitate the communication of measurement rules and results. The most important aspect of measurement is the specification of rules for assigning numbers to characteristics.

b) **Scaling:** Scaling is the assignment of objects to numbers or semantics according to a rule. In scaling, the objects are text statements, usually statements of attitude, opinion, or feeling. For example, consider a scale locating customers of a bank according to the characteristic "agreement to the satisfactory quality of service provided by the branch". Each customer interviewed may respond with a semantic like 'strongly agree', or 'somewhat agree', or 'somewhat disagree', or 'strongly disagree'. We may

even assign each of the responses a number. For example, we may assign strongly agree as '1', agree as '2' disagree as '3', and strongly disagree as '4'. Therefore, each of the respondents may assign 1, 2, 3 or 4.

6.3 ATTITUDE MEASUREMENT TECHNIQUES

Definition of attitude: Attitude has been defined by Gene F. Summers as a Predisposition to respond to an idea or an object. In marketing, this refers to the consumer's predisposition about the product or service. If it is favorable, then the consumer is likely to purchase the product or service. Attitudes about products or services are composed of three elements:

1. **Beliefs** such as the product's strength or the economy of the product or service
2. **Emotional feelings** such as likes or dislikes
3. **Readiness to respond** to the product or service, i.e. to buy it.

Attitude measurement is commonly referred to as ***scaling.***

6.4 NON-DISGUISED AND NON STRUCTURED TECHNIQUES

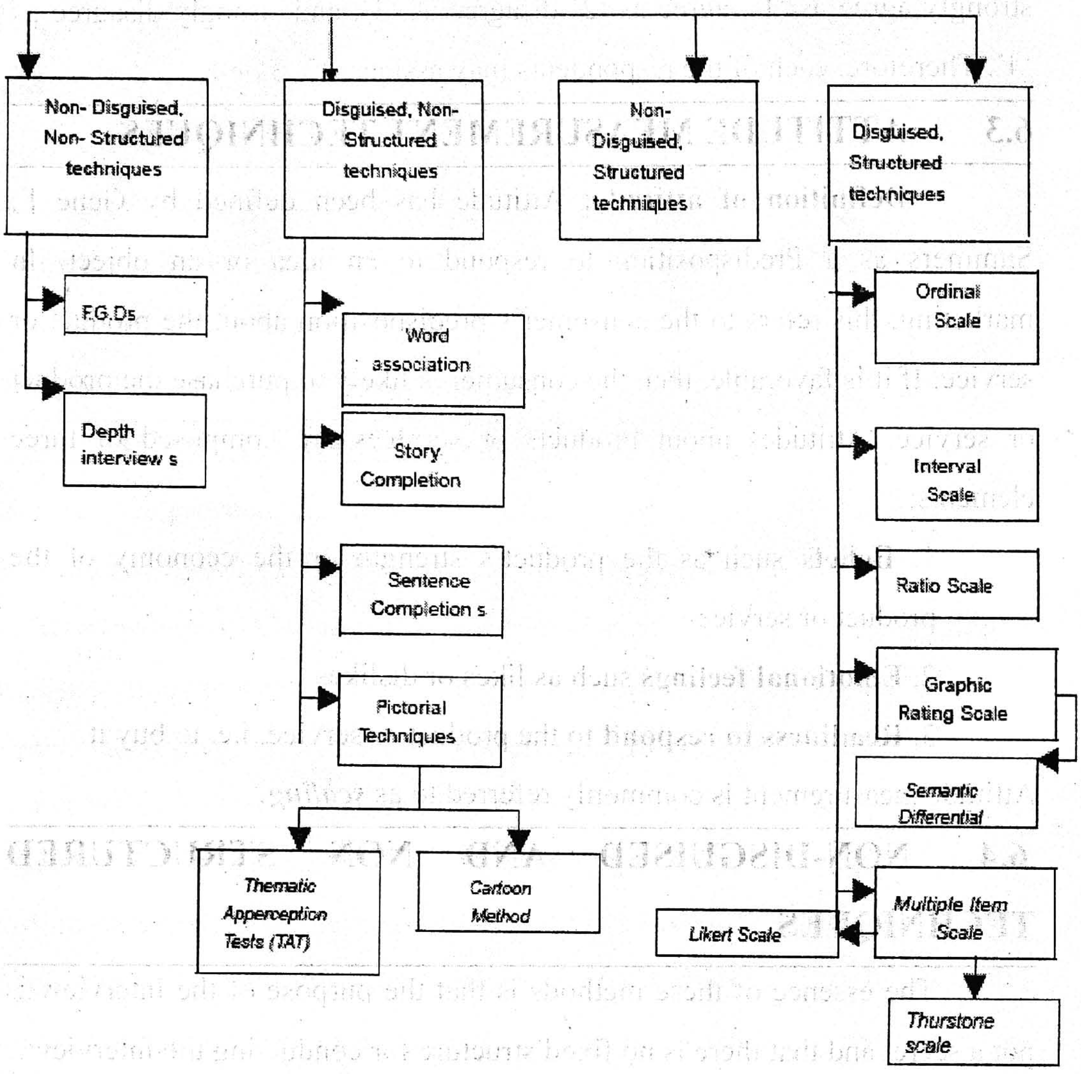
The essence of these methods is that the purpose of the interview is not a secret and that there is no fixed structure for conducting the interview.

6.4.1 Qualitative Research

The most common method of obtaining information about the behavior, attitudes and other characteristics of people is to ask them. However it is not always possible, or desirable to use direct questioning to obtain information. People may be either unwilling or unable to give answers to questions they consider to be an invasion of their privacy, that adversely affect their self-perception or prestige, that are embarrassing that concern motivations that they do not fully understand or cannot verbalize, or for other reasons. Therefore additional approaches to obtaining such information may be necessary.

NOTES

The measurement techniques are divided thus:



Depth interviews and Projective techniques are frequently used by marketing researchers when direct questioning is impractical, more costly, or less accurate. These techniques generally referred to as Qualitative research.

6.4.2 Depth interviews

Check your Progress

1) Is there any specific set of questions to be asked in a depth interview?

Individual depth interviews typically require 30-45minutes. The interviewer does not have a specific set of pre-specified questions that must be asked according to the order imposed by a questionnaire. Instead, there is freedom to create questions, to probe those responses that appear relevant, and generally to try to develop the best set of data in any way practical. However the interviewer must follow one rule; one must not consciously try to affect the content of the answers given by the respondents. The respondent must feel free to reply to the various questions, probes, and other, subtler, ways of

encouraging responses in the manner deemed most appropriate. Subject of interest is discussed in detail.

- There is no fixed pattern for eliciting information from the respondents.
- Generally conducted by highly trained interviewers. They must be thorough in probing the respondents.
- The interviewee is asked about the subject of his choice, coffee, for example, and an attempt is made to explore the respondents' attitudes in depth by probing extensively into any other areas which may come up.
- Interviewers have a general series of topics that they will introduce – perhaps such topics as coffee, or sleep, and will introduce them from time to time if the respondent does not bring them up.
- Tone of the interview is permissive and the respondent is allowed to talk as much as he likes.
- The interviewer must not influence the answers of the respondent.
- The interpretation of the answers is very subjective and knowledge of human behavior is required to analyze the information received.

Individual depth interviews use three questioning techniques namely:

1. ***Laddering*** involves having respondents identify attributes that distinguish brands by asking questions. Each distinguishing attribute is then probed to determine why it is important or meaningful. These reasons are then probed to determine why it is important, and so forth. The purpose is to uncover the "network of meanings" associated with the product, brand, or concept.
2. ***Hidden-issue questioning*** focuses on individual respondents feelings about sensitive issues. Analysis on focus on common underlying themes across respondents. These themes can then be used to guide advertising development
3. ***Symbolic questioning*** requires respondents to describe the opposites of the product/ activity of interest or a specific attribute of the product/ activity.

Individual depth interviews have been found to generate more and higher quality ideas on a per respondent basis than either focus or mini groups. They are particularly appropriate when:

1. Detailed probing of an individual's behavior, attitude or needs is required;
2. The subject matter under discussion is likely to be of a highly confidential nature (e. g. personal investment)
3. The subject matter is of an emotionally charged or embarrassing nature;
4. Certain strong, socially acceptable norms exist (e.g. baby feeding) and the need to conform in a group discussion may influence responses;
5. Where highly detailed understanding of complicated behavior or decision making pattern (e.g. planning the family holiday) are required; or The interviews are with professional people or with people on the subject of their jobs 9 e.g. finance directors)

6.4.3 Focus group discussions (F.G.Ds):

The standard focus group interview involves 8 and 12 individuals and lasts about 2 hours. Normally each group is designed to reflect the characteristics of a particular market segment. The respondents are selected according to the relevant sampling plan and meet at a central location that generally has facility for taping and/ or filming the interviews. In Europe, focus tend to consist of 6 to 8 respondents, vary in length from 1.5 to 4 hours and are often conducted in the home of the recruiter. Otherwise the interviewers are similar. The discussion itself is "led" by a moderator. The moderator attempts to progress through three stages during the interviewer: (1) establish rapport with the group, structure the rules of group interaction, and set objectives; (2) provoke intense discussion in the relevant areas; and (3) summarize the group's responses to determine the extent of agreement. The general either the moderator or a second person prepares a summary of each session after analyzing the session's transcript. Focus Group Interviews can be applied to:

1. Basic- need studies for product idea creation,
2. New product idea or concept exploration,
3. Product positioning studies,
5. Background studies on consumer's frames or reference,

6. Establishment of consumer vocabulary as a preliminary step in questionnaire development and,

7. Determination of attitudes and behavior.

6.4.4 Mini groups

Mini groups consist of a moderator and 4 and 5 respondents rather than the 8 to 12 used in most focus groups. They are used when the issue being investigated requires more extensive probing than is possible in a larger group. Mini groups do not allow the collection of a confidential or highly sensitive data as might be possible in an individual depth interview. However, they do allow the researcher to obtain substantially depth of response on the topics that are covered. Further the intimacy of the small group often allows discussion of quite sensitive issues. The advantages and disadvantages of mini groups are similar to those of standard focus groups, but on a smaller scale.

- In principle, these interviews are the same as the previous ones, excepting that they are conducted in groups rather than for individuals.
- This method is therefore less expensive and less time consuming than the depth interviews.
- This method is advantageous because it gives excellent leads to consumer attitudes that no other method can give.
- Another advantage of this method is that each respondent receives stimulation for responding from his group members and so the interviewer need not prompt the interviewee to answer.
- The disadvantage here is that one or two members could dominate in the group and others might not get a chance to answer. This would again make it an individual effort.

Check your Progress

2) What is a mini group?

6.5 DISGUISED, NON-STRUCTURED TECHNIQUES

The essence of these methods is that the interviewee either does not know that his attitude is being studied, or does not know for which company the survey is being done, or sometimes he does not know both. It involves using various vague stimuli to which the respondent is asked to respond. In

doing so, it is believed that the respondent reveals several elements of his/her attitude that he would not have revealed in the face of direct questions. These tests are not difficult to administer because they are like games played with the respondents. Generally, respondents seem to enjoy the exercise.

6.5.1. Projective Techniques

Projective Techniques are based on the theory that the description of vague objects requires interpretation and this interpretation can only be based on the individual's own background, attitudes, and values. The more vague or ambiguous the object to be described the move one must reveal of oneself in order to complete the description. The following general categories of projective techniques are described: ***association, completion, construction and expression.***

6.5.2. Association Techniques

Association techniques require the subject to respond to the presentation of a stimulus with the first things that come to mind. The word association technique requires the respondent to give the first word or thought that comes to mind after researcher presents a word or phrase. In free association only the first word or thought is required. In successive word association, the respondent is asked to give a series of words or thoughts that occur after hearing a given word. The respondent is generally read a number of relatively neutral terms to establish the technique. Then the words of interest to the researcher are presented, each separated by several neutral terms. The order of presentation of the key words is randomized to prevent any position or order bias from affecting the results. The most common approach to analyzing the resulting data is to analyze the frequency with a particular word or category or word is given in response to the word of interest to the researcher. Word association techniques are used in testing potential brand names and occasionally for measuring attitudes about particular products, product attributes, brands, packages or advertisements.

6.5.3. Completion Techniques

This technique requires the respondent to complete an incomplete stimulus. Two types of completion are of interest to marketing researchers- *sentence completion* and *story completion. Sentence completion,* as the name implies, involves requiring the respondent to complete a sentence. In most sentence completion tests the respondents are asked to complete the sentence with a phrase. Generally they are told to use the first thought that comes to their mind or "anything that makes sense". Because the individual is not required directly to associate himself or herself with the answer conscious or subconscious defenses are more likely to be relaxed and allow a more revealing answer. *Story completion* is an expanded version of sentence completion. As the name suggests part of a story is told and the respondent is asked to complete it.

6.5.4 Construction Techniques

This technique requires the respondent to produce or construct something generally a story, dialogue, or description. They are similar to completion techniques except that less initial structure is provided. *Cartoon techniques* present cartoon-type drawings of one or more people in a particular situation. One or more of the individuals are shown with a sentence in bubble form above their heads and one of the others is shown with a blank bubble that the respondent is to "fill in". Instead of having the bubble show replies or comments, it can be drawn to indicate the unspoken thoughts of one or more of the characters. This device allows the respondent to avoid any restraints that might be felt against having even a carton character *speak* as opposed to *think* certain thoughts. *Third- person techniques* allow the respondent to project attitudes onto some vague third person. This third person is generally "an average woman", "your neighbors", "the guys where you work", "most doctors" or the like. Thus instead of asking the respondent why he or she did something or what he or she thinks about something the researcher asks what friends, neighbors or the average person thinks about the issue. *Picture response,* another useful

construction technique, involves using pictures to elicit stories. These pictures are usually relatively vague, so that the respondent must use his or her imagination to describe what is occurring. *Fantasy scenario* requires the respondent to make up a fantasy about the product or brand. *Personification* asks the respondent to create a personally for the products or brands.

6.5.5 Expressive Techniques

Role-playing is the only expressive technique utilized to any extent by marketing researchers. In *role playing* the consumer is asked to assume the role or behavior of an object or another person, such as a sales representative for a particular department store. The role-playing customer can then be asked to try to sell a given product to a number of different "consumers" who raise varying objections. The means by which the role player attempts to overcome these objections can reveal a great deal about his or her attitudes. Another version of the technique involves studying the role-player's attitudes on what type of people "should" shop at the store in question.

6.5.6 Word Association

One of the oldest and simplest projection techniques. Respondents are presented with a number of different words, one at a time. After each word, they are asked to give the first word that comes to mind. The assumption here is that through free words, the respondents will indicate their inner feelings about the subject. Responses are timed so that those responses that respondents "reason out" are identified and taken into account in the analysis. The time limit is usually 5 seconds. The usual way of constructing such a test is to choose many stimulating and "neutral" words. The words are read out to the respondent one at a time, and the interviewer essentially records the "first word" association by the respondent. Respondents should not be asked to write their responses because then the interviewer will not know if the responses were spontaneous or whether the respondent took time to think out the responses.

6.6.7 Sentence Completion

The respondent is given a number of incomplete sentences and asked to complete them. The rule here too, is that respondent must fill in the first thought that comes to mind. Responses are timed. Here the interviewer gets more information than the word association technique. However, it is difficult to disguise the motive of the study from the respondent, who is usually able to diagnose the investigator's purpose of study. For example, "a man who reads Sportstar is ---------------------------------------." The sentences can be worded in either first or third person. No evidence suggests that one of these approaches could be better than the other.

6.6.8 Story Completion

Respondents are given a half-completed story. This is enough to draw their attention to a particular issue, but the ending is left vague, so that responses can be varied. This technique is very versatile and has numerous applications to marketing problems. The findings about products/ services give companies inputs to determine advertising and promotional themes and product characteristics.

6.6.9 Pictorial Techniques

These are similar to story completion method, except that here pictures are used as the stimuli. The two main methods used here are

a) Thematic Apperception Tests (TAT)

b) Cartoon method

6.6.9.1. Thematic Apperception Tests (TAT)

Clinical psychologists have long used this method. Here the respondent is shown many ambiguous pictures and he is asked to spin stories about them. The interviewer may ask questions to help the respondent to think. For example "what is happening here?" makes the answer focused towards an action. Or "which one is the aggressor?" makes the respondent think about the picture as one of aggression. The reason that respondents must be asked such prompting questions is that the pictures are very abstract and general and as such are open to very broad and irreverent interpretations. So

Check you Progress

3) What is the expansion of TAT?

some amount of focus is needed to channel the respondent's thinking. Each subject in the pictures is a medium through which the respondent projects his feelings, ideas, emotions and attitudes. The respondent attributes these feelings to the characters because he sees in the picture something related to himself. Responses differ widely and analysis depends upon the ambiguity of the picture, the extent to which the respondent is able to guess the conclusions and the vagueness of the support questions asked by the interviewer.

6.6.9.2 Cartoon Tests

They are a version or modification of the TAT, but they are simpler to administer and analyze. Cartoon Characters are shown in a specific situation pertinent to a problem. One or more "*balloons*" indicating the conversation of the characters is left open. The respondent has to then fill these balloons and then analyzed.

6.7 NON- DISGUISED, STRUCTURED TECHNIQUES

The non – structured techniques for attitude measurement are primarily of value in exploratory studies, where the researcher is looking for the salient attributes of given products and the important factors surrounding purchase decisions as seen by the system, one that is more comparable to a scale or yardstick. The term scaling has been applied to the efforts to measure attitudes objectively, and a number of useful scales have been developed.

6.7.1. Nominal Data

A set of data is said to be nominal if the values / observations belonging to it can be assigned a code in the form of a number where the numbers are simply labels. You can count but not order or measure nominal data. Nominal Scale is the crudest among all measurement scales but it is also the simplest scale. In this scale the different scores on a measurement simply indicate different categories. The nominal scale does not express any values or relationships between variables. For example, labeling men as '1' and women as '2' which is the most common way of labeling gender for data

Check your Progress

4) Cartoon test is the modification of which other projective test?

recording purpose does not mean women are 'twice something or other' than men. Nor it suggests that men are somehow 'better' than women. Another example of nominal scale is to classify the respondent's income into three groups: the highest income as group 1. The middle income as group 2, and the low-income as group 3. The nominal scale is often referred to as a categorical scale. The assigned numbers have no arithmetic properties and act only as labels. The only statistical operation that can be performed on nominal scales is a frequency count. We cannot determine an average except mode. In designing and developing a questionnaire, it is important that the response categories must include all possible responses. In order to have an exhaustive number of responses, you might have to include a category such as 'others', 'uncertain', 'don't know', or 'can't remember' so that the respondents will not distort their information by forcing their responses in one of the categories provided. Also, you should be careful and be sure that the categories provided are mutually exclusive so that they do not overlap or get duplicated in any way.

6.7.2. Ordinal Scales

Ordinal Scale involves the ranking of items along the continuum of the characteristic being scaled. In this scale, the items are classified according to whether they have more or less of a characteristic. For example, you may wish to ask the TV viewers to rank the TV channels according to their preference and the responses may look like this as given below:

TV Channel	Viewers preferences
Doordarshan-I	1
Star plus	2
NDTV News	3
Aaaj Tak TV	4

The main characteristic of the ordinal scale is that the categories have a logical or ordered relationship. This type of scale permits the measurement of degrees of difference, (that is, 'more' or 'less') but not the specific amount of differences (that is, how much 'more' or less'). This scale is very common in

marketing, satisfaction and attitudinal research. Another example is that a fast food home delivery shop may wish to ask its customers:

How would you rate the service of our staff?
(1) Excellent • (2) Very Good • (3) Good • (4) Poor • (5) Worst

. Here we rank the attitudes, opinions and preferences from best to worst or from worst to best. However, the amount of difference between the ranks cannot be found out. Using ordinal scale data, we can perform statistical analysis like Median and Mode, but not the Mean. They are the simplest attitude measuring scales used in marketing research. They serve to rank respondents according to some characteristics such as favorabiliy to a certain brand, or to rank items such as brands in order of consumer preference. They do not measure the degree of favorability of the different rankings. All the scale tells is that the individual or item has more, less, or the same amount of the characteristic being measured as some other time. They are the most widely used type of scales in marketing research. A set of data is said to be ordinal if the values / observations belonging to it can be ranked (put in order) or have a rating scale attached. You can count and order, but not measure, ordinal data. The categories for an ordinal set of data have a natural order., for example, suppose a group of people were asked to taste varieties of biscuit and classify each biscuit on a rating scale of 1 to 5, representing strongly dislike, dislike, neutral, like, strongly like. A rating of 5 indicates more enjoyment than a rating of 4, for example, so such data are ordinal. However, the distinction between neighboring points on the scale is not necessarily always the same. For instance, the difference in enjoyment expressed by giving a rating of 2 rather than 1 might be much less than the difference in enjoyment expressed by giving a rating of 4 rather than 3.

6.7.3. Interval Scales

Interval Scale is a scale in which the numbers are used to rank attributes such that numerically equal distances on the scale represent equal distance in the characteristic being measured. An interval scale contains all the

information of an ordinal scale, but it also one allows to compare the difference/distance between attributes. For example, the difference between '1' and '2' is equal to the difference between '3' and '4'. Further, the difference between '2' and '4' is twice the difference between '1' and '2'. However, in an interval scale, the zero point is arbitrary and is not true zero. This, of course, has implications for the type of data manipulation and analysis. We can carry out on data collected in this form. It is possible to add or subtract a constant to all of the scale values without affecting the form of the scale but one cannot multiply or divide the values. Measuring temperature is an example of interval scale. We cannot say 40^0C is twice as hot as 20^0C. The reason for this is that ^{0}C does not mean that there is no temperature, but a relative point on the Centigrade Scale. Due to lack of an absolute zero point, the interval scale does not allow the conclusion that 40^0C is twice as hot as 20^0C. Interval scales may be either in numeric or semantic formats. The following are two more examples of interval scales one in numeric format and another in semantic format.

Interval Scale in Numeric Format

Food supplied is:						*Indicate your score on the concerned blank and circle the appropriate number on each line.*
Fresh	1	2	3	4	5	
Tastes good	1	2	3	4	5	
Value for money	1	2	3	4	5	
Attractive packaging	1	2	3	4	5	
Prompt time delivery	1	2	3	4	5	

Interval Scale in Semantic Format

Food supplied is:	Excellent	Very Good	Good	Poor	Worst
Fresh					
Tastes good					
Value for money					
Attractive packaging					
Prompt time delivery					

The interval scales allow the calculation of averages like Mean, Median and Mode and dispersion like Range and Standard Deviation. They separate individuals or items by rank order but measure the distance between rank positions in equal units. Such a scale permits the researcher to say that the position 4 is above position 3 on the scale, and also the distance from position 5 to 4 is same as from 4 to 3. Such a scale however does not permit conclusions that position 6 is twice as strong as position 3 because no zero position has been established. An interval scale is a scale of measurement where the distance between any two adjacent units of measurement (or 'intervals') is the same but the zero point is arbitrary. Scores on an interval scale can be added and subtracted but cannot be meaningfully multiplied or divided. Examples of interval scales include the heights of tides, and the measurement of longitude.

6.7.4. Ratio Scales

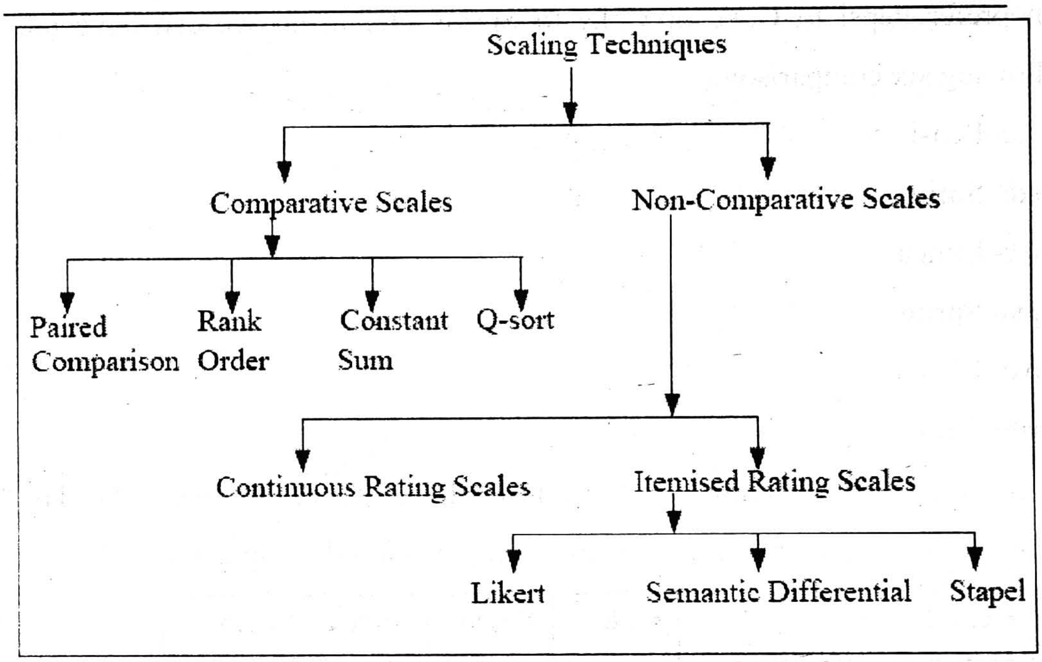
Ratio Scale is the highest level of measurement scales. This has the properties of an interval scale together with a fixed (absolute) zero point. The absolute zero point allows us to construct a meaningful ratio. Examples of ratio scales include weights, lengths and times. In the marketing research, most counts are ratio scales. For example, the number of customers of a bank's ATM in the last three months is a ratio scale. This is because you can compare this with previous three months. Ratio scales permit the researcher to compare both differences in scores and relative magnitude of scores. For example, the difference between 10 and 15 minutes is the same as the difference between 25 and 30 minutes and 30 minutes is twice as long as 15 minutes. Most financial research that deals with rupee values utilizes ratio scales. However, for most behavioural research, interval scales are typically the highest form of measurement. Most statistical data analysis procedures do not distinguish between the interval and ratio properties of the measurement scales and it is sufficient to say that all the statistical operations that can be performed on interval scale can also be performed on ratio scales. If one measures the distance between two points as four feet

Check your Progress

5) What are the main difference between interval cale and the atio scale in terms of their Properties

and between two other points as two feet, it is possible say that one distance is twice that of the other because each distance is measured from an absolute zero. A scale that permits such measurements is called ratio scale. While ratio scales are common in physical science, the measurement of attitudes is still so crude that they are of little significance in marketing research. Now you must be wondering why you should know the level of measurement. Knowing the level of measurement helps you to decide on how to interpret the data.

6.8 TYPES OF SCALING TECHNIQUES



The various types of scaling techniques used in research can be classified into two categories: (a) comparative scales, and (b) Non-comparative scales. In **comparative scaling**, the respondent is asked to compare one object with another. For example, the researcher can ask the respondents whether they prefer brand A or brand B of a detergent. On the other hand. in non comparative scaling respondents need only evaluate a single object. Their evaluation is independent of the other object which the researcher is studying. Respondents using a non-comparative scale employ whatever rating standard seems appropriate to them. Non-comparative techniques consist of continuous and itemized rating scales. Figure below shows the classification of these scaling techniques.

6.8.1 Comparative Scales

The comparative scales can further be divided into the following four types of scaling techniques: (a) Paired Comparison Scale, (b) Rank Order Scale, (c) Constant Sum Scale, and (d) Q-sort Scale.

6.8.1.a) Paired Comparison Scale: This is a comparative scaling technique in which a respondent is presented with two objects at a time and asked to select one object (rate between two objects at a time) according to some criterion. The data obtained are ordinal in nature. For example, there are four types of cold drinks - Coke, Pepsi, Sprite, and Limca. The respondents can prefer Pepsi to Coke or Coke to Sprite, etc. In all we can have the following six comparisons.

Coke–Pepsi

Coke–Sprite

Coke–Limca

Pepsi–Sprite

Pepsi–Limca

Sprite–Limca

In general, with n brands we have n(n −1)/ 2 paired comparisons. The following is the data recording format using the paired comparisons.

Brand	Coke	Pepsi	Sprite	Limca
Coke	—	Ö		
Pepsi		—		
Sprite	Ö	Ö	—	
Limca	Ö	Ö	Ö	—
No. of times preferred	2	3	1	0

A Ö in a particular box means that the brand in that column was preferred over the brand in the corresponding row. In the above recording, Coke was preferred over Sprite, Coke over Limca, in this case the number of times coke preferred was 2 times. Similarly, Pepsi over Coke, Pepsi over Sprite, Pepsi over Limca, in this case Pepsi was 3 time preferred. Thus, the number

of times a brand was preferred is obtained by summing the Ö s in each column. The following table gives paired comparison of data (assumed) for four brands of cold drinks.

Brand	Coke	Pepsi	Sprite	Limca
Coke	–	0.90	0.64	0.14
Pepsi	0.10	–	0.32	0.02
Sprite	0.36	0.68	–	0.15
Limca	0.86	0.98	0.85	–

The entries in the boxes represent the proportion of respondents preferring 'column brand' and to 'row' brand. For example, 90% prefer Pepsi to Coke and only 10% prefer Coke to Pepsi, etc. Paired comparison is useful when the number of brands is limited, since it requires direct comparison and overt choice. One of the disadvantages of paired comparison scale is violation of the assumption of transitivity may occur. For example, the respondent preferred Coke 2 times, Pepsi 3 times, Sprite 1 time, and Limca 0 times. That means, preference-wise, Pepsi >Coke, Coke >Sprite, and Sprite > Limca. However, the number of times Sprite was preferred should not be that of Coke. In other words, if A>B and B >C then C >A should not be possible. Also, the order in which the objects are presented may bias the results. The number of items/brands for comparison should not be too many. As the number of items increases, the number of comparisons increases geometrically. If the number of comparisons is too large, the respondents may become fatigued and no longer be able to carefully discriminate among them. The other limitation of paired comparison is that this scale has little resemblance to the market situation, which involves selection from multiple alternatives. Also, respondents may prefer one item over certain others, but they may not like it in an absolute sense.

6.8.1.b) Rank Order Scale: This is another type of comparative scaling technique in which respondents are presented with several items

simultaneously and asked to rank them in the order of priority. This is an ordinal scale that describes the favoured and unfavoured objects, but does not reveal the distance between the objects. For example, if you are interested in ranking the preference of some selected brands of cold drinks, you may use the following format for recording the responses. Like paired comparison, the rank order scale, is also comparative in nature. The resultant data in rank order is ordinal data. This method is more realistic in obtaining the responses and it yields better results when direct comparisons are required between the given objects. The major disadvantage of this technique is that only ordinal data can be generated.

Preference of cold drink brands using rank order scaling

Instructions: Rank the following brands of cold drinks in order of preference. Begin by picking out the one brand you like most and assign it a number1. Then find the second most preferred brand and assign it a number 2. Continue this procedure until you have ranked all the brands of cold drinks in order of preference. The least preferred brand should be assigned a rank of 4. Also remember no two brands receive the same rank order.

Format:

Brand	Rank
(a) Coke	3
(b) Pepsi	1
(c) Limca	2
(d) Sprite	4

6.8.1. c) Constant Sum Scale: In this scale, the respondents are asked to allocate a constant sum of units such as points, rupees, or chips among a set of stimulus objects with respect to some criterion. For example, you may wish to determine how important the attributes of price, fragrance, packaging, cleaning power, and lather of a detergent are to consumers. Respondents might be asked to divide a constant sum to indicate the relative importance of the attributes using the following format. "If an attribute is assigned a higher number of points, it would indicate that the attribute is more important." From the above Table, the price of the detergent is the

most important attribute for the consumers followed by cleaning power, packaging. Fragrance and lather are the two attributes that the consumers cared about the least but preferred equally." The advantage of this technique is saving time.

Importance of detergent attributes using a constant sum scale

Instructions: Between attributes of detergent please allocate 100 points among the attributes so that your allocation reflects the relative importance you attach to each attribute. The more points an attribute receives, the more important the attribute is. If an attribute is not at all important, assign it zero points. If an attribute is twice as important as some other attribute, it should receive twice as many points.

Format:

Attribute	**Number of Points**
(a) Price	50
(b) Fragrance	05
(c) Packaging	10
(d) Cleaning Power	30
(e) Lather	05
Total Points	100

However, there are two main disadvantages. The respondents may allocate more or fewer points than those specified. The second problem is rounding off error if too few attributes are used and the use of a large number of attributes may be too taxing on the respondent and cause confusion and fatigue.

6.8.1. d) **Q-Sort Scale:** This is a comparative scale that uses a rank order procedure to sort objects based on similarity with respect to some criterion. The important characteristic of this methodology is that it is more important to make comparisons among different responses of a respondent than the responses between different respondents. Therefore, it is a comparative method of scaling rather than an absolute rating scale. In this method the respondent is given statements in a large number for describing the characteristics of a product or a large number of brands of a product. For example, you may wish to determine the preference from among a large number of magazines. The following format shown in

Table below may be given to a respondent to obtain the preferences. Preference of Magazines is obtained by using Q-Sort Scale Procedure. Note that the number of responses to be sorted should not be less than 60 or not more than 140. A reasonable range is 60 to 90 responses that result in a normal or quasi-normal distribution. This method is faster and less tedious than paired comparison measures. It also forces the subject to conform to quotas at each point of scale so as to yield a quasi-normal distribution. The utility of Q-sort in marketing research is to derive clusters of individuals who display similar preferences, thus representing unique market segments.

Instructions: The bag given to you contain pictures of 90 magazines. Please choose 10 magazines you 'prefer most', 20 magazines you 'like', 30 magazines to which you are 'neutral (neither like nor dislike)', 20 magazines you 'dislike', and 10 magazines you 'prefer least'. Please list the sorted magazine names in the respective columns of the form provided to you.

Format:

Prefer Most	Like	Neutral	Dislike	Prefer Least
(10)	(20)	(30)	(20)	(10)

6.8.2 Non-Comparative Scales

The non-comparative scaling techniques can be further divided into: (a) Continuous Rating Scale, and (b) Itemised Rating Scale.

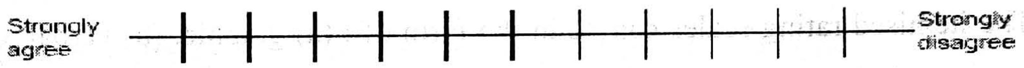
6.8.2 a) Continuous Rating Scales

It is very simple and highly useful. In continuous rating scale, the respondent's rate the objects by placing a mark at the appropriate position

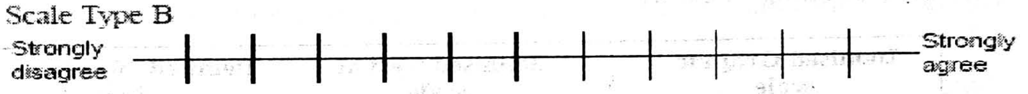
on a continuous line that runs from one extreme of the criterion variable to the other. Examples of continuous rating scale are given below:

Question: How would you rate the TV advertisement as a guide for buying?

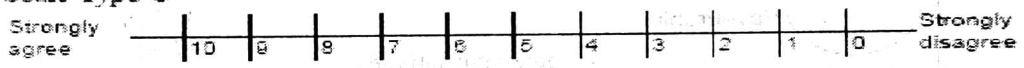
Scale Type A



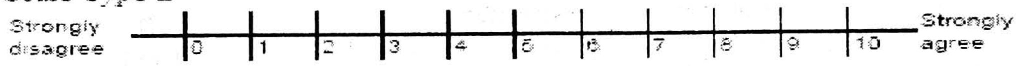
Scale Type B



Scale Type C



Scale Type D



When scale type A and B are used, the respondents score is determined either by dividing the line into as many categories as desired and assigning the respondent a score based on the category into which his/her mark falls, or by measuring distance, in millimeters, centimeters, or inches from either end of the scale. Which ever of the above continuous scale is used, the results are normally analysed as interval scaled.

6.8.2 b) Itemised Rating Scales

Itemised rating scale is a scale having numbers or brief descriptions associated with each category. The categories are ordered in terms of scale position and the respondents are required to select one of the limited number of categories that best describes the product, brand, company, or product attribute being rated. Itemised rating scales are widely used in marketing research. Some rating scales may have only two response categories such as : agree and disagree. Inclusion of more response categories provides the respondent more flexibility in the rating task. Consider the following questions:

1. How often do you visit the supermarket located in your area of residence?

• Never, • Rarely, • Sometimes, • Often, • Very often

2. In your case how important is the price of brand X shoes when you buy them?

- Very important, • Fairly important, • Neutral, • Not so important

The itemised rating scales can be in the form of : (a) graphic, (b) verbal, or (c) numeric as shown below:

Itemised Graphic Scale	Itemised Verbal Scale	Itemised Numeric Scale
Favourable	Completely satisfied Somewhat satisfied	−5 −4 −3 −2
Indifferent	Neither satisfied nor dissatisifed	−1 0 +1
	Somewhat dissatisfied	+2 +3
Unfavourable	Completely dissatisfied	+4 +5

In this section we will discuss three itemised rating scales, namely (a) Likert scale, (b) Semantic Differential Scale, and (c) Stapel Scale.

a) **Likert Scale**

These scales are sometimes referred to as summated scales. It requires a respondent to indicate a degree of agreement or disagreement with each of a series of statements related to the attitude object.

For Example: The service at a retail store is very important to me:

____ Strongly Agree ____ Agree ____ Neither Agree nor Disagree ____ Disagree ____ Strongly Disagree

To analyze a Likert Scale, each response category is assigned a numerical value. These examples could be assigned values such as Strongly Agree=1, through Strongly Disagree=5 or the scoring could be reversed., or a –2 through +2 system could be used. They can be analyzed on an item-by-item basis, or they can be summed to form a single score for each individual.

Advantages

1. It is relatively easy to construct and administer.

2. Instructions that accompany the scale are easily understood; hence it can be used for mail surveys and interviews with children.

Disadvantages

1. It takes a longer time to complete as compared to Semantic Differential Scales, etc.

2. Care needs to be taken when using Likert Scales in cross cultural research, as there may be cultural variations in willingness to express disagreement.

In business research, the Likert scale, developed by Rensis Likert, is extremely popular for measuring attitudes, because, the method is simple to administer. With the Likert scale, the respondents indicate their own attitudes by checking how strongly they agree or disagree with carefully worded statements that range from very positive to very negative towards the attitudinal object. Respondents generally choose from five alternatives (say strongly agree, agree, neither agree nor disagree, disagree, strongly disagree). Consider the following example of a study or measuring attitudes towards cricket.

	Strongly agree	Agree	Not sure	Disagree	Strongly disagree
It is more fun to play a tough, competitive cricket match than to play an easy one.	5	4	3	2	1

To measure the attitude, the researchers assign weights or scores to the alternative responses. In the above example the scores 5 to 1 are assigned to the responses. Strong agreement of the respondent indicates the most favourable attitudes on the statement, and the score 5 is assigned to it. On the other hand, strong disagreement of the respondent indicates the most unfavourable attitude on the statement, and the score 1 is assigned to it. If a negative statement towards the object is given, the corresponding scores would be reversed. In this case, the response 'strongly agree' will get a score

of 1 and the response 'strongly disagree' will get a score of 5. A Likert scale may include a number of items or statements. Each statement is assumed to represent an aspect of an attitudinal domain. For example, Table below shows the items in a Likert Scale to measure opinions on food products.

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
If the price of raw materials fall, firms too should reduce the price of the food products.	1	2	3	4	5
There should be uniform price through out the country for food products	1	2	3	4	5
The food companies should concentrate more on keeping hygiene while manufacturing food products.	1	2	3	4	5
The expiry dates should be printed on the food products before they are delivered to consumers in the market.	1	2	3	4	5
There should be government regulations on the firms in keeping acceptable quality and on the prices	1	2	3	4	5
Now-a-days most food companies are concerned only with profit making rather than taking care of quality.	1	2	3	4	5

b) Semantic Differential Scale

This is a seven point rating scale with end points associated with bipolar labels (such as good and bad, complex and simple) that have semantic meaning. The Semantic Differential scale is used for a variety of purposes. It can be used to find whether a respondent has a positive or negative attitude towards an object. It has been widely used in comparing brands, products and company images. It has also been used to develop advertising and promotion strategies and in a new product development study. Look at the following Table, for examples of Semantic Differential Scale.

NOTES

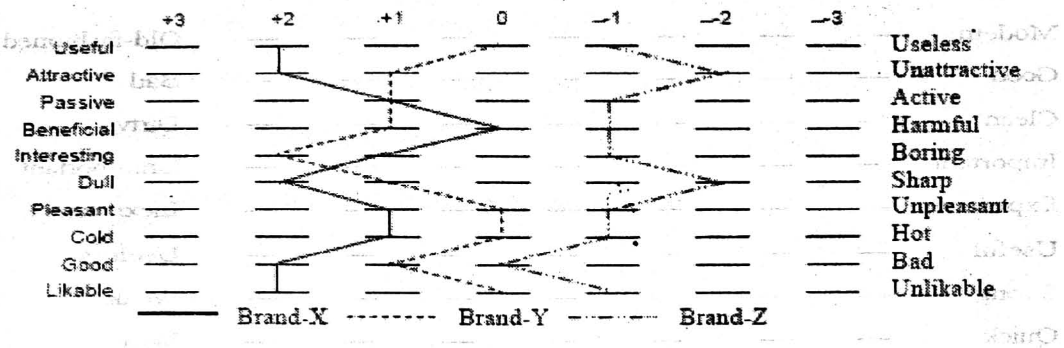
Modern	—	—	—	—	—	—	—	**Old-fashioned**
Good	—	—	—	—	—	—	—	**Bad**
Clean	—	—	—	—	—	—	—	**Dirty**
Important	—	—	—	—	—	—	—	**Unimportant**
Expensive	—	—	—	—	—	—	—	**Inexpensive**
Useful	—	—	—	—	—	—	—	**Useless**
Strong	—	—	—	—	—	—	—	**Weak**
Quick	—	—	—	—	—	—	—	**Slow**

In the Semantic Differential scale only extremes have names. The extreme points represent the bipolar adjectives with the central category representing the neutral position. The in between categories have blank spaces. A weight is assigned to each position on the scale. The weights can be such as +3, +2, +1, 0, –1, –2, –3 or 7, 6, 5, 4, 3, 2, 1. The following is an example of Semantic Differential Scale to study the experience of using a particular brand of body lotion.

In my experience, the use of body lotion of Brand-X was:

	+3	+2	+1	0	–1	–2	–3	
Useful	—	—	—	—	—	—	—	Useless
Attractive	—	—	—	—	—	—	—	Unattractive
Passive	—	—	—	—	—	—	—	Active
Beneficial	—	—	—	—	—	—	—	Harmful
Interesting	—	—	—	—	—	—	—	Boring
Dull	—	—	—	—	—	—	—	Sharp
Pleasant	—	—	—	—	—	—	—	Unpleasant
Cold	—	—	—	—	—	—	—	Hot
Good	—	—	—	—	—	—	—	Bad
Likable	—	—	—	—	—	—	—	Unlikable

In the semantic Differential scale, the phrases used to describe the object form a basis for attitude formation in the form of positive and negative phrases. The negative phrase is sometimes put on the left side of the scale and sometimes on the right side. This is done to prevent a respondent with a positive attitude from simply checking the left side and a respondent with a negative attitude checking on the right side without reading the description of the words. The respondents are asked to check the individual cells depending on the attitude. Then one could arrive at the average scores for comparisons of different objects. The following Figure shows the experiences of 100 consumers on 3 brands of body lotion.



In the above example, first the individual respondent scores for each dimension are obtained and then the average scores of all 100 respondents. for each dimension and for each brand were plotted graphically. The maximum score possible for each brand is + 30 and the minimum score possible for each brand is –30. Brand-X has score +14. Brand-Y has score +7, and Brand-Z has score –11. From the scale we can identify which phrase needs improvement for each Brand. For example, Brand-X needs to be improved upon benefits and Brand-Y on pleasantness, coldness and likeability. Brand Z needs to be improved on all the attributes.

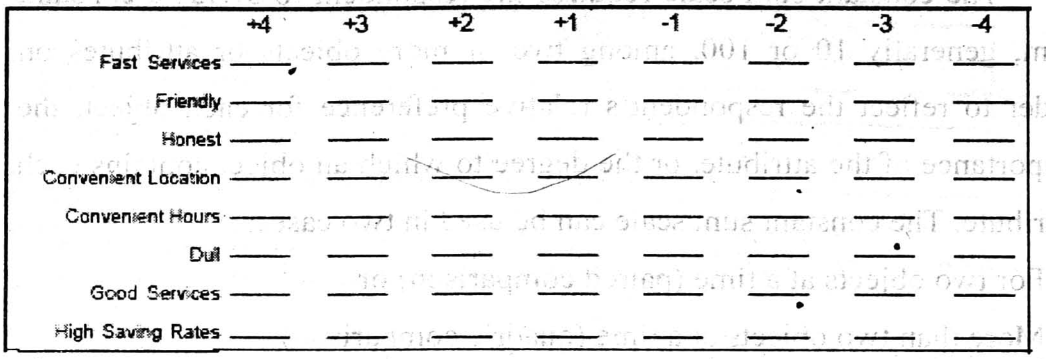
c) Staple Scale: The Stapel scale was originally developed to measure the direction and intensity of an attitude simultaneously. Modern versions of the Stapel scale place a single adjective as a substitute for the Semantic differential when it is difficult to create pairs of bipolar adjectives. The modified Stapel scale places a single adjective in the centre of an even number of numerical values (say, +3, +2, +1, 0, –1, –2, –3). This scale measures how close to or how distant from the adjective a given stimulus is perceived to be. The following is an example of a Staple scale.

Instructions: Select a plus number for words that you think describe personnel banking of a bank accurately. The more accurately you think the word describes the bank, the larger the plus number you should choose. Select a minus number for words you think do not describe the bank accurately. The less accurately you think the word describes the bank, the larger the minus number you should choose.

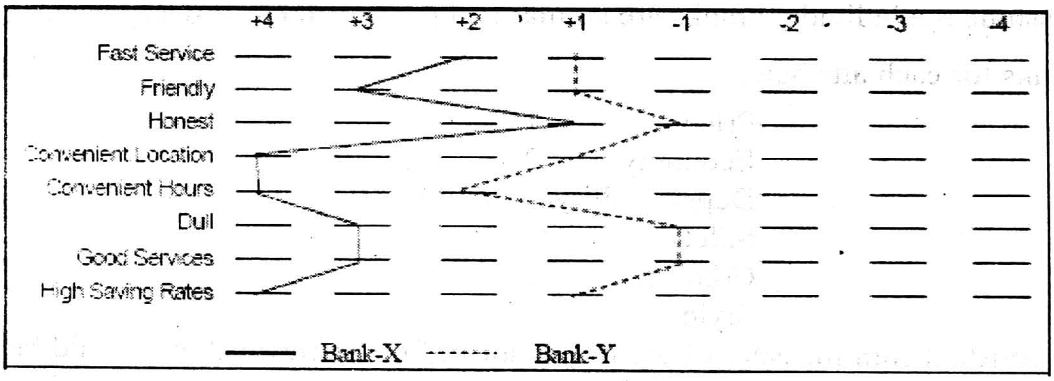
Format:

+5	+5
+4	+4
+3	+3
+2	+2
+1	+1
Friendly Personnel	**Competitive Loan Rates**
–1	–1
–2	–2
–3	–3
–4	–4
–5	–5

The following format shows an example of Stapel scale that Illustrates respondents description on personnel banking of a bank.



Each respondent is asked to circle his opinion on a score against each phrase that describes the object. The final score of the respondent on a scale is the sum of their ratings for all the items. Also, the average score for each phrase is obtained by totaling the final score of all the respondents for that phrase divided by the number of respondents of the phrase. The following Figure shows the opinions of 100 respondents on two banks.



In the above example first the individual respondent's scores for each phrase that describes the selected bank are obtained and then the average scores of all 100 respondents for each phrase are plotted graphically. The maximum score possible for each bank is +32 and the minimum possible score for each brand is –32. In the example, Bank-X has score +24, and Bank-Y has score +3. From the scale we can identify which phrase needs improvement for each Bank.

6.9 THE CONSTANT SUM SCALE

The constant sum scale requires the respondent to divide a constant sum, generally 10 or 100, among two or more objects or attributes on order to reflect the respondent's relative preference for each object, the importance of the attribute, or the degree to which an object contains each attribute. The constant sum scale can be used in two cases:

1. For two objects at a time (paired comparison) or
2. More than two objects at a time (quadric comparison)

6.9.1 Advantages

When rank order data is used the researcher has no way of knowing the characteristic which is of overwhelming importance or of knowing a characteristic which is not much more important than other characteristics. This can be explained with the following example:

Suppose a sample of respondents from the target market is requested to rank order several automobile characteristics with 1 being more important. Assume the individual ranks are similar and produce the following median ranks for each attribute:

Price	1
Economy	2
Dependability	3
Safety	4
Comfort	5
Style	6

A constant sum measure of the importance of the same attributes could be obtained from the following procedure: Divide 100 points among the characteristics listed so that the division will reflect how important each characteristic is to your selection of a new automobile.

Price	____
Economy	____
Dependability	____
Safety	____
Comfort	____
Style	____
Total	100

All three of the following groups' average responses to the constant sum scale would be consistent with the rank order results just described:

	Group A	Group B	Group C
Price	35	20	65
Economy	30	18	9
Dependability	20	17	8
Safety	10	16	7
Comfort	3	15	6
Style	2	14	5
	100	100	100

With rank order scale the researcher has no way of knowing if price is of importance (GROUP C); part of a general, strong concern for overall cost (GROUP A); or not much important than the other attributes (GROUP B). Constant Sum Scale provides such evidence.

6.9.2 Disadvantage

A disadvantage could be that individuals could occasionally misassign points such that the total is more than, or less than 100. This can be adjusted for by dividing each point allocation by the actual total and multiplying the result by 100.

6.10 THURSTONE SCALE

It is one of the *Multi Item Scales* developed by L.L. Thurstone's method of *Equal Appearing Intervals* on the concept that, even though people could not assign quantitative measures to their own attitudes, they could tell the difference between the attitude represented by two different statements and could identify items that were approximately halfway between the two. The procedure is as follows

1. Collect a large number of statements (perhaps as may as several hundred) related to the attitude in question
2. Have a number of judges (perhaps 20 or more) sort the statements independently into 11 piles that vary from the most favorable statement to neutral statements to most unfavorable statements.

3. Study the frequency distribution of ratings for each statement and eliminate those statements that the different judges have given widely scattered ratings – that are in a number of different piles

4. Determine the scale value of each of the remaining statements – that is, the number of the pile in pile in which the median of the distribution falls

5. Select one of the two statements from each of the 11 piles for the final scale. Those statements with the narrowest range of rating are preferred as the most reliable.

6.10.1. Advantages

It is important to note that there are 11 attitude positions because in a scale with odd number of parameters, it is easier to identify a neutral position.

6.10.2. Disadvantages

1. Thurstone scales are not widely used for Marketing Research because the are time consuming during preparation 2. The ratings may be influenced by the Judges' personal attitude

3. Different individuals can obtain exactly the same score from agreeing with quite different items.

4. It does not obtain information about the intensity of agreement with the ratings

6.10.3. Comparison of Thurstone and Likert Scale

It is obvious that these two scales have a lot in common. They have been widely used in the past. Due to the ordinal nature of the Likert scales, many individuals feel they it may be more reliable that the Thurstone Scale.

6.11 DISGUISED, STRUCTURED TECHNIQUES

The basis premise underlying such tests is that the respondents will reveal their attitudes by the extent to which their answers to the objective questions vary from the correct answers to objective questions that vary from the correct answers. Respondents are provided with questions that they are no able to answer correctly. Thus, they are forced to guess at the answers. The direction and extent of these guessing errors is assumed to reveal their

attitudes on the subject. For example, individuals tend to gather information that supports their attitudes and, therefore, the extent and kind of information individuals possess on a given subject indicate something of their attitude. For example:

- How much do u think it cost for the hot cereal alone in a average bowl of cereal such as you'd serve at the breakfast?
- Do corn flakes cost less or more per bowl than cereal?

Selection of an appropriate scaling technique

In this unit, so far, you have learnt some of the important scaling techniques that are frequently used in attitudinal research for the measurement of attitudes. Each of these techniques has some advantages and disadvantages. Now you may ask which technique is more appropriate to use to measure attitudes. Virtually any technique can be used to measure the attitudes. But at the same time all techniques are not suitable for all purposes. As a general rule, you should use a scaling technique that will yield the highest level of information feasible in a given situation. Also, if possible the technique should permit you the use of a variety of statistical analysis. A number of issues decide the choice of scaling technique. Some significant issues are:

1) **Problem Definition and Statistical Analysis:** The Choice between ranking, sorting, or rating techniques is determined by the problem definition and the type of statistical analysis likely to be performed. For example, ranking provides only ordinal data that limits the use of statistical techniques.

2) **The Choice between Comparative and Non-comparative Scales:** Some times it is better to use a comparative scale rather than a non-comparative scale. Consider the following example: How satisfied you are with the brand- X detergent that you are presently using? Completely

Somewhat	***Neither***	***Somewhat***	***Completely***	
satisfied	***satisfied***	***satisfied nor dissatisifed***	***dissatisfied***	***dissatisfied***

This is a non-comparative scale since it deals with a single concept (the brand of a detergent). On the other hand, a comparative scale asks a respon... to rate a concept. For example, you may ask: Which one of the following brands of detergent you prefer?

Brand-X ***Brand-Y***

In this example you are comparing one brand of detergent with another brand. Therefore, in many situations, comparative scaling presents 'the ideal situation' as a reference for comparison with actual situation.

3) **Type of Category Labels:** We have discussed different types of category labels used in constructing measurement scales such as verbal categories and numeric categories. Many researchers use verbal categories since they believe that these categories are understood well by the respondents.

4) **Number of Categories:** While there is no single, optimal number of categories, traditional guidelines suggest that there should be between five and nine categories. Also, if a neutral or indifferent scale response is possible for at least some of the respondents, an odd number of categories should be used.

5) **Balanced versus Unbalanced Scale:** In general, the scale sh... be balanced to obtain objective data.

6) **Forced versus Non-forced Categories:** In situations where the respondents are expected to have no opinion, the accuracy of data may be improved by a non forced scale that provides a 'no opinion' category.

6.12 SUMMARY

Thus we can see that there are four levels of measurements: nominal, ordinal, interval, and ratio and the measurement scales, commonly used in marketing research, can be divided into two types; comparative and non-comparative scales.

6.13 ANSWERS TO 'CHECK YOUR PROGRESS'

1) There are no pre established questions for a dept interview.

2) Mini groups consist of a moderator and 4 and 5 respondents rather than the 8 to 12 used in most focus groups. They are used when the issue being investigated requires more extensive probing than is possible in a larger group.

3) Thematic Apperception Tests (TAT)

4) Cartoon test is the modification of Thematic Apperception Tests (TAT)

5) The main difference between interval scale and the ratio scale in terms of their properties Interval scale does not have a fixed (absolute) zero point whereas ratio scale has a fixed zero point that allows us to construct a meaningful ratio.

6.14 EXERCISES AND QUESTIONS

1) What are the different levels of measurement? Explain any two of them.

2) How do you select an appropriate scaling technique for a research study? Explain the issues involved in it.

3) Discuss briefly the issues involved in attitude measurement.

4) Differentiate between ranking scales and rating scales. Which one of these scales is better for measuring attitudes?

5) In what type of situation is the Q-sort technique more appropriate?

6) Construct a Rank Order Scale to measure toothpaste preferences. Discuss its advantages and disadvantages.

7) Construct a Semantic differential scale to measure the experiences of respondents in using Brand-X of shaving cream (assume that all the respondents use that brand).

6.15 FURTHER READINGS

1.Aaker, David A. and George S. Day. (1983) Marketing Research, John Wiley, New York.

2.Donald S. Tull and Gerald S. Albaum. (1973) Survey Research: A Decisional

3.Meister, David. (1985) Behavioural Analysis and Measurement Methods, John Wiley, New York.

NOTES

UNIT VII

DATA PROCESSING

STRUCTURE

7.1 Introduction

7.2 Data processing

7.3 Preliminary screening of the data collected

7.4 Data Editing

7.5 Data coding

7.6 classifications of data

7.7 Tabulation of data

7.8 Processing of data

7.9 Summary

7.10 Answers to 'Check Your Progress'

7.11 Exercises and Questions

7.12 Further Readings

UNIT OBJECTIVES

After reading this unit you would be able to

- Understand the basic concepts of data processing
- To understand the procedural aspects of editing
- To understand the steps in coding

- To classify data
- To tabulate the data
- Know the basics of data analysis

7.1 INTRODUCTION

Once the data has been collected it has to be processed, analyzed and interpreted. The efforts put in by the researcher and his team will go waste in case the information collected is not properly processed and analyzed. Data processing is to do with the specialized activities of editing, coding, classification and tabulation of the collected data so that they are amenable to analysis and finally interpretation.

7.2 DATA PROCESSING

Technically speaking, data processing means and includes the activities of editing, coding, classification and tabulation of collected data to make them amenable to analysis and interpretation. In fact, data processing means reduction of data to such an extent that one is able to under analysis and latter on the interpretation. Data as collected are unduly large which have to be boiled down to a definite size and shape. It means taking out unwanted data and correcting it to make it more relevant and meaningful. The processing of data is a type of office work which can be attended to by the office staff under the guidance of researcher. The steps involved in the processing of data are as noted below:

1. Preliminary screening of the data collected,
2. Editing of data,
3. Coding of data,
4. Classification of data,
5. Tabulation of data,
6. Analysis of data.

Check your Progress

1) What are the major steps in data processing?

NOTES

7.3 PRELIMINARY SCREENING OF THE DATA COLLECTED

Preliminary screening or scanning of data is prior to the editing. The purpose of initial screening is to verify accuracy and completeness of data for further processing. During the data collection process, supervisors keep check on the data collection work of investigators. Thereafter, the preliminary screening is undertaken at the office level. The screening is usually on the basis of the following three points:

(i) Relevance: It is a fact that major portion of data collected through questionnaire is useful for classification. However, certain information may or may not be useful for actual survey work. For example, investigators may note the views, opinions or other details based on their observation during the interview. Such information needs screening in order to judge its relevance to the problem under study. This enables the researcher to use relevant information while drawing conclusions.

(ii) Reliability: Data used for MR should be reliable, complete and correct. In the preliminary screening of data, some more tests are applied to check the extent of reliability of data. An attempt is being made to verify whether adequate care has been taken for the collection

of reliable' data in the initial stages. Data of doubtful quality are discarded or sent back for further verification.

(iii) Practical use / utility: Investigators are usually instructed to note additional information that seems pertinent to the research problem. As a result, additional information is noted by them in the questionnaires. Some extra information may be relevant but of little practical value. Such information need not be considered for classification. It is possible to omit such data from further processing work.

NOTES

7.4 EDITING OF DATA

The first task in data processing is the editing. Editing is the process of eliminating errors in the raw data collected so that the data are prepared for subsequent coding. As it is a very subjective process, it is necessary that persons, who are well-qualified and trained in the job of editing, should alone be entrusted with this responsibility.

In all cases where editorial corrections are to be made, it is necessary that these should be kept distinct from the changes made either by the respondent or by the interviewer. This can be ensured by the editor by using a different coloured pencil for editing the raw data. An editor should be well-versed with the editing of questionnaires. It may be emphasised that editing a manuscript is different from the editing of .a questionnaire or numeric data. People who are good at editing descriptive material may not be able to edit numeric data satisfactorily.

Persons who are quite efficient in detecting flaws or errors in the data in just one glance should be entrusted with this job. Enumerators with long experience and having a special aptitude for editing of data should be given preference over others. When the services of more than one editor are required, it is advisable to give each one exploit editing instructions in order to ensure consistency in the editing of data.

7.4.1 Types of data editing:

Editing can be undertaken both at the time when the field survey is in progress and when it has been completed. Based on this there are two types of editing of data.

1. Field Editing,

2. Office Editing.

Check your Progress

2) What is editing?

7.4.1.1 Field Editing

Field editing means editing of data during the process of field survey itself. When the interviewer fills in the information at the time of the interview, he often uses several abbreviations due to the paucity of time. These need to be spelt out fully later. It is advisable for the interviewer to carefully look into the questionnaire at the earliest possible opportunity after the interview so that he can enter the proper responses or complete the partial answers.

7.4.1.2 Office Editing

Office editing (also called Central Editing) is possible when all questionnaires are collected at the head office for tabulation. As far as possible, a single editor should carry out this task so that consistency in editing can be ensured. Office editing is more systematic and uniform in character as compared to field editing. In the office editing, errors in answers, fictitious interviews, and inconsistencies in answers by respondents, vagueness in answers, irrelevant answers or incorrect answers are rectified or omitted from further processing. There are two ways of editing raw data. These are: (a) Manual Editing, and (b) Computer Editing. Computer editing is better as it is accurate, faster and economical. Manual editing, on the other hand, is expensive, time-consuming and also provides scope for human errors.

7.4.2 Procedure of Editing:

Editing of data is a skilled job and persons doing editing work must be trained, experienced and sincere. They should be given suitable instructions for the editing work. Persons doing editing work should use different colour pencils for editing work as compared to the colour used for answering the questionnaire. As a result, the original answers and the edited answers are clearly visible. In addition, the editing procedure should be laid down in writing. Detailed instructions about the method of dealing with

different types of answers should be given to the editing staff. This brings speed, uniformity and accuracy in the data editing work. Editors are usually supposed to deal with five types of answers recorded by the interviewers but in a uniform way. Such answers are:

(i) The incorrect answer: If the mistake in the answer is minor, the editor may 'edit' it. However, if the answer is false to a significant extent, the editor has to delete the entire answer.

(ii) The incomplete answer: Very little editing work is possible if the answer is incomplete. If the answer is related to other question in the questionnaire, the missing data may be filled in on the basis of the information given in other questions. Incomplete answers should be completed by the editors only when they are very sure as to what the respondents desire to communicate.

(iii) The inconsistent answer: Sometimes, answers to two closely related questions are not consistent with each other. Here, the editor should try to find out the real intention of the respondent. The editor has to use his judgment to decide which answer is correct and adjust all other answers accordingly. He can treat one answer as correct and adjust the other answers accordingly. This brings consistency due to editing.

(iv) The *"Don't* know" answer or "No reply" answer: If such replies are limited and insignificant from the point of view of survey work, they should be ignored. If they are large in number, then, one of the following three options can be used:

(i) Distribute the "don't know" randomly among the other categories.

(ii) Show the "don't know" or "No reply" as a regular category in addition to the existing ones.

(iii)Estimate the possible answers from other data contained in the questionnaire.

NOTES

(v) The answer not available in standard form: In this case, the respondent gives answer to the question but not in the specific unit of measurement. The quantity/units noted in the answer are different from the units specified. Here, the editor has to standardize the answer. Such edited answer should be treated as complete and correct answer.

7.5 CODING OF DATA

Coding is the procedure of classifying the answers to a question into meaningful categories. It helps in arranging/grouping of varied responses into manageable categories for the purpose of classification, tabulation and analysis of data. In the absence of coding, it will not be possible to reduce a large number of heterogeneous responses into meaningful categories. The symbols used to indicate these categories are called codes. For example, the letter M may be n for male and F for female. Similarly, numbers like 1, 2, 3 can be used as code numbers. In large surveys, where mostly structured questionnaires are used, the response categories are pre-determined and are contained in the questionnaires themselves. If the questionnaire itself bears code numbers for each type of answer, it is called a pre-coded questionnaire. Pre-coding gives certain advantages like speed, accuracy and legibility. However, pre-coding is not feasible when all possible answers cannot be anticipated / accounted for. Here, the space can be provided for assigning code numbers to such questions. Some examples of pre-coded questions are given below:

Questions	*Answers*	*Codes*
How often these days do you go to the cinema?	More than once a week	1
	Once a week	2
	Once a fortnight	3
	Once a month	4
	Three or four times a year	5
	Less often	6
	Never	7
Which type of wrist watch do you own?	Hand wound	1
	Automatic	2
	Electronic	3
Which of the following battery operated equipment do you have in your family?	Torch	1
	Transistor	2
	Others (specify)	3

Check your Progress

3) What is coding?

Sometimes data are transferred from the questionnaire to a coding sheet and then punched into cards. Coding may be done simultaneously along with editing or it may be done after editing. Frequently followed practice is to edit and code the data simultaneously. These two operations can be treated as one operation and can be looked after by one person alone. It is possible to do coding after editing and also giving these jobs to two persons.

7.5.1 Guidelines for Data Coding

The first step is to specify the different categories/classes into which the responses are to be classified. The second step is to allocate individual answers to different categories. Attention should be given to the following points while coding the data available in the questionnaire

(1) Coding should cover all types of answers given by the respondents. It should be of wide range. For this, it is desirable to add a category with 'other replies'.

(2) Coding categories should not overlap. The categories introduced should be all inclusive and mutually exclusive. Every answer should legitimately fit into only one category and not in many categories.

(3) A code sheet containing the code number of all categories of answers for each question in the questionnaire should be prepared for quick reference to those who are concerned with coding work.

(4) Coding work should be given to a knowledgeable person so that mistakes in coding will be avoided and the quality of data for further processing will be superior.

(5) Appropriate use of numbers, symbols and alphabets should be made for coding of information given in the questionnaire.

(6) More stress should be given to pre-coding over post-coding.

7.5.2 Editing Vs Coding Of Data

	Editing of Data	Coding of Data
Meaning	Is the process of eliminating errors in the raw data collected so that the data may subsequently undergo tabulation quickly and correctly	Is the process of translating respondents' replies into numerical form / codes for the purpose of easy and quick tabulation
Timing	Editing is possible during the field survey and after the collection of data through field survey	Coding is possible at the time of preparing the questionnaire, along with editing of data or after the collection of data
Purpose	To check the accuracy and consistency in the recording of responses	To have quick and correct tabulation of data collected
Necessity	Editing is necessary even when the data for is limited in volume	Coding of data is necessary particularly when the data to be classified are large in volume

7.6 CLASSIFICATION OF DATA

Classification of data is next to coding. Data collected are not directly useful for tabulation. For this the classification of data is essential. The aim of classification is to arrange data systematically for quick and easy understanding / analysis.

7.6.1 Objectives of Classification of Data

(1) To condense the available data in a compact manner for quick understanding and also for comparison.

(2) To make the data ready for further processing and analysis.

(3) To bring out important facts and also to facilitate comparison with the data of similar nature.

(4) To make data ready for tabulation.

(5) To facilitate easy and quick interpretation of data.

7.6.2 Basic Principles in Classification of Data

(1) **Classification should not be ambiguous:** The very purpose of classification is to remove ambiguity. For this, homogeneity should be made the basis of classification. Various categories/groups should be clearly defined in classification. This avoids confusion in classification and facilitates analysis of concrete nature.

(2) **Should be on the basis of a single classification principle.** The consideration for classification should be only one at a time. For example, buyers may be classified on the basis of economic or social status or frequency of purchases or the amount spent. For classification, only one point at a time should be taken into account. If more than one point needs to be applied, separate category should be created for each point.

(3) **Should be mutually exclusive:** This means the answer given by a respondent should be suitable for placement under one classification only. There should be no overlapping of the categories after classification.

(4) **Should be mutually exhaustive:** Classification should cover all data under one category or the other. For this all possible categories should be prepared at the time of classification. Answers of the respondents should not be omitted in the process of classification.

(5) **Should be action-oriented:** Classification should be devised keeping in view the future action to be taken on the basis of the analysis and interpretation of data. Classification is a means and not the end in itself.

Check your Progress

4) What is classify- - cation?

NOTES

(6) **Should be distinct:** Various sub-divisions under each category of classification should be 'so distinct as to indicate differences substantial enough to lead to different decisions to suit each specific aspect of the problem.

(7) **Should be pertinent to Marketing Research project:** The purpose of classification should be to study the marketing problem in the best possible manner. Classification without reference to the purpose will not be useful for analysis of data.

7.6.2 Methods of Classification of Data

Classification of data is possible on the following basis

(1) **Geographical Classification:** In geographical classification, the available data are divided on geographical basis i.e. area-wise or region-wise. Geographical classification is quite simple, convenient and used extensively For example: sale of products of a company may be classified on state basis. Here, the territory/area covered by one state is taken as a base for classification. In census reports, for example, tables show composition of population in relation to different states.

(2) **Chronological/Periodical Classification:** In chronological classification, the data are arranged/classified as per the time or period. Here, year or month is used as a base for classification. For example, in census reports, population increase is shown period wise and ten years period is taken as a base.

(3) **Qualitative Classification:** Here, the available data are classified on the basis of some quality, attributes or qualifications. For example, population can be classified as males and females or literates and illiterates or rural and urban and so on. This classification is also called "classification by attributes". A retail trader can classify his regular customers on the basis of their monthly purchases.

(4) **Quantitative Classification:** In quantitative classification, the available data are classified on actual quantitative measurement. This method is also known as "classification by variables". Variable is a characteristic which changes/varies from observation to observation. Retail shopkeepers for example, can be classified on the basis of their monthly turnover.

7.7 TABULATION OF DATA

Tabulation is next to editing, coding and classification. It prepares quantitative data in an understandable and concise form. Tabulation means the sorting of data, allocating the same to relevant classification categories and counting the number of cases that belong to each category. It involves noting of information in the assembly sheets and preparation of summary tables for quick understanding. The simplest way to tabulate is to count the number of responses to one question. This is also called "univariate tabulation". The analysis based on just one variable is obviously meager. Where two or more variables are involved in tabulation, it is called "bivariate or multivariate tabulation". In marketing research projects generally both types of tabulation are used.

The tabulation may be done by hand or by machine or some part by hand and the other by machine. The number of tabulations will depend on the number of variables, while the number of responses to a question will depend on the sample size. If the number of variables in a survey is not large and the sample size is small, then tabulation by hand is preferable. On the other hand, when the number of variables involved is large as a result of which there may be several relationships and the sample size is large, it will be more appropriate to opt for machine tabulation.

7.7.1 Features of Tabulation of Data

(i) In tabulation, the data are displayed in the form of statistical tables.

(ii) Tabulation is an orderly grouping of data or arrangement of data in rows and columns, which facilitates comparisons, contrasts and crosschecking of data.

(iii)It aims at systematically presenting the data to elucidate the problem to be solved through marketing research.

(iv)Tabulation facilitates analysis and interpretation of data for drawing conclusions. In the absence of tabulation, analysis of data may be difficult and defective.

(v) Tabulation converts bare statistics into meaningful statistical tables for analysis.

(vi)Tabulation involves two stages. In the first stage, the tabulators record the information found in the questionnaire in the form called "assembly sheet" or "counting sheet" or "tabulation sheet". Such tabulation sheets are divided into columns for noting different categories of responses. In the second stage, summary tables are prepared which provide certain conclusions for analysis.

(vii) Tabulation work needs to be conducted in an orderly manner and for this the tabulation procedure should be decided clearly preferably in writing.

(viii) It is a time-consuming and costly operation and is also susceptible to possibilities of errors. For avoiding such errors, the tabulation work should be given to a team of experienced and skilled tabulators. In addition, the detailed procedure for tabulation should be prepared for avoiding confusion and errors on the part of tabulators.

7.7.2 Need of Tabulation

(1) Tabulation is needed as it brings orderly arrangement of data for analysis. It acts as a connecting link between organisation of data and the interpretation of data.

(2) Tabulation of data is needed as it brings accuracy and speed in arranging data for drawing conclusions. It provides data in the form of statistical tables.

(3) Tabulation of data is needed as it facilitates correct analysis and interpretation of data. It can be used in order to find out the correlation between two or more variables.

(4) Tabulation is needed as a stepping-stone to the analysis and interpretation of data. The analysis of data may be difficult and even defective if the tabulation of data is not proper.

(5) Tabulation of data is needed as it enables a researcher to draw conclusions properly from the collected data and to present those conclusions clearly in the form of report and recommendations.

(6) Tabulation is absolutely needed for the analysis of data collected. It is a statistical device useful for arranging data into meaningful statistical tables for analysis and for drawing conclusions. Tabulation provides information in a summary form. In brief, tabulation of data is needed for arranging the data in a compact manner for easy and quick analysis and interpretation.

7.7.3 Steps in tabulation of data:

(1) **Programming:** Programming means preparing proper planning for tabulation of collected data. Here, the available information is broken down in the order of tabulation, For this, the actual tables are prepared. Thereafter, schedule is prepared. This schedule shows the groups into which the questionnaires are to be sorted out and the order in which the work is to be carried out. Thus, programming means complete planning of the tabulation work to be undertaken.

(2) **Numbering of questionnaires:** In this second stage, all questionnaires (duly edited and coded) to be used in tabulation work are serially numbered. This facilitates accuracy in the tabulation work. In addition, easy reference from cards to source information is also possible. Numbering should be done before starting actual tabulation work. Here, every respondent is given a specific number and also to the material (i.e. information) supplied by him.

(3) **Sorting of questionnaires:**

After numbering the questionnaires, they should be sorted out as per the relevant groups. Sorting is useful for quick and easy counting for tabulation purpose. This provides the base count to the tabulators. Base count indicates the total number of respondents who have sent complete replies and naturally their questionnaires are not rejected. After the base counts are finalised, they are noted in a table showing the base counts of each major classification category. This facilitates effective check on the tabulation work.

(4) **Counting:** After sorting questionnaires, the next step is the counting of individual answers. The usual method of counting is putting four paralleled strokes with the fifth stroke cutting these four. (E.g. IIII). The groups of five ticks each can be counted quickly and conveniently. The counting work can be divided amongst the tabulators on the basis of area or classification categories. After counting, the results are noted on the counting sheets. Such sheets are separately prepared for each category of classification.

(5) **Verifying the counting work:** Verification of tabulation work is essential particularly when many tabulators work for different classification categories. The chief tabulator has to share this responsibility. Numbering of forms, coding and tabulation cards must be verified for complete accuracy in the tabulation work.

(6) **Summarizing**:

In the summarizing step, various totals and subtotals obtained in the counting will be computed on a number of tally sheets. The totals brought together on summary sheets are useful for the construction of tables. The tables show information in a compact manner. All conclusions drawn should be checked twice to eliminate errors in the counting work.

NOTES

7.7.4 Importance of Tabulation of Data

Tabulation is an important activity in marketing research process. It provides information in a statistical summary for easy understanding by the researcher during the process of analysis and interpretation. It facilitates analysis of data with a view to establishing relationship and highlighting significant differences in the information collected through questionnaires. The analysis of data will be defective, if the tabulation of data is not proper. Tabulation facilitates presentation of data in concise and understandable form. There is saving of time and space. Tabulation enables a researcher to draw conclusions in a clear and concise manner. It creates convenient framework for the analysis of data and report writing. Tabulated data are also incorporated in the final report to support the conclusions and recommendations. According to **J. R. Hicks,** "Tabulated and classified data speaks themselves whereas disorganised data are like dead meat."

7.7.5 Methods of tabulation of data:

There are two methods used for the tabulation of data.

(a) Manual/Hand Tabulation,

(b) Machine/Mechanical/Electronic Tabulation.

The suitability of a method will depend on the nature of the survey and the speed with which the findings are needed.

(a) Manual/Hand Tabulation

In manual tabulation, the sorting out of questionnaires and their numbering is done by hands only. This sorting is as per the major classifications fixed. The counting is noted on the counting sheets in sets of five tally marks. Hand tabulation is used when a small sample survey is undertaken and only few households are interviewed and a few questionnaires (upto 500) are collected.

Hand/Manual tabulation was used extensively in the olden days. This traditional method is not convenient for national level surveys and when quick analysis of data is required. For example, a survey is conducted as regards movies seen by selected respondents (sample) within one week. For a sample of say, 60 respondents, the tabulation may be as shown below

No of movies seen				*Frequency*
0	卌	IIII		9
1	卌	卌	卌	15
2	卌	卌	III	13
3	卌	卌		10
4	卌	III		8
5 or more	卌			5
				60

The hand tally can also be used for preparing cross-classification tables Consider, for example, the relationship between the economic status and the number of movies seen.

No. of movies seen	*Rich*	*Middle*	*Poor*	*Total*
0	—	IIII	卌	9
1	—	卌 II	卌 III	15
2	II	IIII	卌 II	13
3	III	III	IIII	10
4	I	I	卌 I	8
5 or more	IIII	I	—	5
	10	20	30	60

Merits of hand/manual tabulation:

(1) **Simplicity:** Hand tabulation is simple. Mechanical devices including card-punch machines and computers are not needed. Only work sheets are adequate for tabulation work.

(2) **Economy:** Hand tabulation is not expensive as one or two people are adequate for tabulation of limited number of questionnaires.

(3) **Coding, etc. are not required:** In hand tabulation, coding of the data and other preparatory work are not required.

(4) **Convenient:** Hand tabulation is convenient when the forms to be tabulated are few and the tables to be prepared are simple.

(5) **Limited errors:** Errors are intermittent in hand tabulation and do not continue as in the case of faulty computer program.

(6) **Experts are not required:** The services of experts are not required? for manual/hand tabulation.

(7) **Suitable for the study of complicated problem:** Hand tabulation is convenient when marketing problem under study is complicated and numerous replies are reported by respondents.

Demerits of hand/manual tabulation:

(1) **Time-consuming and costly:** Hand tabulation is often time-consuming and costly particularly when the volume of tabulation work is substantial and complicated in character.

(2) **Limited speed:** The speed of hand tabulation is limited as compared to mechanical tabulation.

(3) **Not suitable for large number of questionnaires:** It is not suitable when large numbers of questionnaires are to be tabulated.

(4) Detailed tables are not possible: In hand tabulation, large number of detailed tables and cross-tabulations are not possible.

(5) Not suitable to present computer age: Hand tabulation is not convenient method in the present world of computers.

(6) Exposed to clerical/human errors: The chances of clerical and human errors are more in hand/manual tabulation as the entire tabulation process relies on human hands and memory.

(b) Machine/Mechanical Tabulation

In the case of large scale surveys, where a good deal of data from a large number of respondents has been collected, hand tabulation will be bothersome and take much longer to complete. In such a case, the method of machine tabulation is used. Here, sorting, classification and tabulating data are done with the help of machines/mechanical devices. Such machines include electronic calculators, tabulators and now computers. The data are first coded and transferred to punch cards by punching holes for further processing. The cards are mechanically sorted out as per classifications, counted and printed in tabular form. A wide-variety of tabulating machines including computers, are available for this purpose. Large organisations purchase such machines for tabulation work. Small organisations can get such machines on hire basis.

Mechanical tabulation is useful for organising the tabulation work with speed and accuracy. The use of computers in data processing has increased at a fantastic rate during the past decade. It is rightly said that the first industrial revolution freed man's muscles while the second is freeing his mind for more challenging pursuits. Computer processing is particularly convenient and economical when large volume of input is available and also when greater speed and accuracy are needed in the data processing. Computers handle complex data speedily and accurately. Very many details are available easily and clearly because of the use of computers.

Merits of machine/mechanical tabulation:

(1) Speed and economy: Machine tabulation is faster and quicker and naturally the results are available quickly. It is also economical particularly when the volume of survey work is substantial. Extensive surveys can be handled easily with machine tabulation.

(2) Cross-tabulation and cross classification possible: Machine tabulation permits the introduction of various types of cross-tabulation and cross-classification. Thus, more detailed information is available accurately and quickly. This raises the efficiency of interpretation and decision-making.

(3) Accuracy: Machine tabulation is always accurate provided adequate care is taken while feeding data. Human factor is eliminated and naturally there is complete accuracy in mechanical tabulation.

(4) Re-tabulation easy: In machine tabulation, re-tabulation is possible easily, quickly and also at low cost.

(5) Integration of many operations: In machine tabulation, wide variety of operations can be performed and complex mathematical techniques can be used extensively and accurately.

(6) In-built corrections possible: Some machines are fitted with built-in machine errors checking devices. This facilitates automatic corrections of mistakes.

(7) Convenient for large volume of work: Machine tabulation is always convenient when large numbers of questionnaires are to be tabulated (census data). It is a must when the data for tabulation are voluminous and varied in character. For repeated studies, machine tabulation is convenient.

(8) Avoids monotony: Monotony and drudgery of tabulation can be avoided by passing on tabulation work to the machines.

Demerits of Machine/Mechanical Tabulation:

(1) **Special arrangements necessary:** Machine tabulation requires additional coding and other arrangements of data for tabulation.

(2) **Expensive:** It is expensive as costly machines and trained staffs are required for tabulation work. This is particularly true when the tabulation work is limited and simple in nature.

(3) **Mechanical errors possible:** In mechanical tabulation, errors may continue due to computer programming flaws. Mistakes are also possible due to feeding of wrong information.

(4) **Leads to confusion during interpretation of data:** Machine tabulation gives wide variety of information to researcher and this may create confusion at the time of interpretation of data.

(5) **Longer lead-time:** Machine tabulation needs longer lead-time particularly when computers are to be used for tabulation work.

(6) Rigidity: Machine tabulation in rather rigid as machines operates as per the orders issued.

7.7.6 Meaning and Components of a Table

Tabulation provides information in a compact manner. This facilitates quick understanding of the data. It is also useful for interpretation of data. In tabulation, different types of tables are prepared. A table contains information under rows and columns for quick understanding and interpretation. This gives convenience to readers of the research report. Summary tables are usually included in the final research report for easy understanding of the report and the conclusions drawn. An ideal table has different components as explained below:

(1) **Title:** Every table needs clear and concise title, which suggests the details regarding the data with their collection and classification. The title

speaks about the background of the data presented in the table. It gives convenience to the reader of the table.

(2) **Head note:** Head note is written just below the title in a bracket. It indicates the unit in which the data are presented in the table. For example: Rs. in lakhs/millions or Weight in tones. Head note gives convenience while analysing the table and also while drawing conclusions by the researcher.

(3) **Stubs:** The stubs indicate the row headings or row titles. They are usually placed at the extreme left hand column.

(4) **Caption:** Caption is the heading/title given to a column. It states clearly what each column represents. It is a heading given to a column. A table may have one or two captions. The caption should be simple and clear so as to give clear understanding to the reader.

(5) **Sub-entries:** Sub-entries indicate the sub-headings under the stubs and show what each row represents.

(6) **Body:** Body represents the main part of the table. It contains statistical data as per the other indications (like caption and sub-entries) of the table. The statistical data need to be studied in the light of other details given in the table. A table is made up of vertical columns with horizontal rows.

(7) **Footnote:** A footnote is given if some specific information about a particular item needs to be communicated to the readers. The item concerned with the footnote is marked with an asterisk and exactly similar asterisk mark is placed where the footnote is given. The footnote is normally given at the bottom of the table.

(8) **Source:** Sometimes, the table is reproduced from another magazine, census report yearbook and so on. In the case of such tables, source of information should be mentioned. The source includes the name and date of publication and other details.

Check your Progress

5) What is a stub?

7.7.7 Kinds/types of tables:

(1) **Primary Tables:** Primary tables or counting sheets are the sheets on which information is classified into different categories on the basis of certain considerations. In such a table, available data are grouped under different heads for counting purpose. Such tables are useful for arranging data for analysis of significant relationships and differences. Primary tables are not directly useful for drawing specific conclusions but only for tabulating available information in a systematic manner.

No. of movies seen	*Rich*	*Middle*	*Poor*	*Total*
0	—	IIII	卌	9
1	—	卌 II	卌 III	15
2	II	IIII	卌 II	13
3	III	III	IIII	10
4	I	I	卌 I	8
5 or more	IIII	I	—	5
	10	20	30	60

(2) **Summary Tables:** Summary tables are tables which summarise the results of counting. They are useful for drawing definite conclusions as they provide data in a compact form for analysis and interpretation. In marketing research, primary tables are necessary for organising the data collected while summary tables are necessary and useful for analysis and interpretation of data. Summary tables are prepared by using primary tables as base. Summary tables may be further classified into three major groups.

(a) Simple / One-dimensional / Univariate Tables.

(b) Two-dimensional / Bivariate Tables.

(c) Complex / Multivariate Tables.

(a) **Simple / One-dimensional / Univariate Tables:** One-dimensional tables are simple tables with only one variable. Such tables are simple to prepare but are not of much use for analysis of data. They provide limited

information. A table showing number of persons using different brands of washing soap or a table indicating classification of workers in a factory as per monthly wage payment are the examples of simple/one dimensional tables. E.g. Movies seen by Sample Respondents

No of movies seen per week	Number of respondents	Percentage of respondents
0	9	15
1	15	25
2	13	22
3	10	17
4	8	13
5	5	8
	60	100

(b) **Two-dimensional / Bivariate Tables:** Two-dimensional table correlates two variables. They provide information about two inter-related characteristics of the subject under study. For example, a table showing monthly salary of workers in a factory can be further classified on the basis of sex (sex-wise). Such tables give more information and are used extensively in marketing research. Two-dimensional tables are also called cross tabulation tables. Such tables bring out causal relationships and also highlight the factors having a major influence on the data. Cross tabulation provide information in a compact manner and casual relationship between certain factors. Two-dimensional tables are superior to one dimensional or simple table. E.g. Preference for Shopping Centre by Income Level of Households

Income level	Shopping centre		Total no. of households
	A	B	
Up to Rs 1000	100	200	300
Rs 1000+	120	80	200
Total	220	280	500

NOTES

(c) **Complex / Multivariate Tables:** Tables showing more than two Data Processing classifications are called complex tables. Such tables may be three or four dimensional in character. For example, monthly salary of workers in a factory may be classified on the basis of amount, sex, age and number of dependents. Such complex tables provide a large amount of information in a compact form and also help in bringing out major relationships. However, such tables may be difficult to understand easily and quickly. E.g. Purchase behavior by income and age

Age	35-50		35-30		16-34		16-34	
Income	High		Low		High		Low	
	n	%	n	%	n	%	n	%
Product A	85	85	48	80	10	20	27	18
Purchase	15	15	12	20	40	80	123	82
	100	100	60	100	50	100	150	100

7.7.8 Generally Accepted Principles of Tabulation

Such principles of tabulation, particularly of constructing statistical tables, can be briefly stated as follows:

1. Every table should have a clear, concise and adequate title so as to make the table intelligible without reference to the text and this title should always be placed just above the body of the table.

2. Every table should be given a distinct number to facilitate easy reference.

3. The column headings (captions) and the row headings (stubs) of the table should be clear and brief.

4. The units of measurement under each heading or sub-heading must always be indicated.

5. Explanatory footnotes, if any, concerning the table should be placed directly beneath the table, along with the reference symbols in the table.

6. Source or sources from where the data in the table have been obtained must be indicated just below the table.

7. Usually the columns are separated from one another by lines, which make the table more readable and attractive. Lines are always drawn at the top and bottom of the table and below the captions.

8. There should be thick lines to separate the data under one class from the data under another class and the lines separating the sub-divisions of the classes should be comparatively thin lines.

9. The columns may be numbered to facilitate reference.

10. Those columns whose data are to be compared should be kept side by side. Similarly, percentages and/or averages must also be kept close to the data.

11. It is generally considered better to approximate figures before tabulation as the same would reduce unnecessary details in the table itself.

12. In order to emphasize the relative significance of certain categories, different kinds of type, spacing and indentations may be used.

13. It is important that all column figures be properly aligned. The decimal points and (+) or (-) signs should be in perfect alignment.

14. Abbreviations should be avoided to the extent possible and ditto marks should not be used in the table.

15. Miscellaneous and exceptional items, if any, should be usually placed in the last row of the table.

16. Table should be made as logical, clear, accurate and simple as possible. If the data happen to be very large, they should not be crowded in a single table for that would make the table unwieldy and inconvenient.

17. Total of rows should normally be placed in the extreme right column and that of columns should be placed at the bottom.

18. The arrangement of the categories in a table may be chronological, geographical, alphabetical or according to magnitude to facilitate comparison. Above all, the table must suit the needs and requirements of an investigation.

7.8 PROCESSING OF DATA

We can take up the following two problems of processing the data for analytical purposes:

(a) **The problem concerning "Don't know" (or DK) responses:** While processing the data, the researcher often comes across some responses that arc difficult to handle. One category of such responses may be 'Don't Know Response' or simply DK. response. When the DK response group is small, it is of little significance. But when it is relatively big, it becomes a matter of major concern in which case the question arises: Is the question which elicited DK response useless? The answer depends on two point's viz.; the respondent actually may not know the answer or the researcher may fail in obtaining the appropriate information. In the first case the concerned question is said to be alright and DK response is taken as legitimate DK response. But in the second case, DK response is more likely to be a failure of the questioning process.

How DK responses are to be dealt with by researchers? The best way is to design better type of questions. Good rapport of interviewers with respondents will result in minimizing DK responses. But what about the DK responses that have already taken place? One way to tackle this issue is to estimate the allocation of DK answers from other data in the questionnaire. The other way is to keep DK responses as a separate category in tabulation where we can consider it as a separate reply category if DK responses happen to be legitimate, otherwise we should let

the reader make his own decision. Yet another way is to assume that DK responses occur more or less randomly and as such we may distribute them among the other answers in the ratio in which the latter have occurred. Similar result will be achieved if all DK replies are excluded from tabulation and that too without inflating the actual number of other responses.

(b) **Use of percentages:** Percentages are often used in data presentation for they simplify numbers, reducing all of them to a 0 to 100 range. Through the use of percentages, the data are reduced in the standard form with base equal to 100 which fact facilitates relative comparisons. While using percentages, the following rules should be kept in view by researchers:

1. Two or more percentages must not be averaged unless each is weighted by the group size from which it has been derived.

2. Use of too large percentages should be avoided, since a large percentage is difficult to understand and tends to confuse, defeating the very purpose for which percentages are used.

3. Percentages hide the base from which they have been computed. If this is not kept in view, the real differences may not be correctly read.

4. Percentage decreases can never exceed 100 per cent and as such for calculating the percentage of decrease, the higher figure should invariably be taken as the base.

5. Percentages should generally be worked out in the direction of the causal-factor in case of two-dimension tables and for this purpose we must select the more significant factor out of the two given factors as the causal factor.

7.8.1 Types of Analysis

As stated earlier, by analysis we mean the computation of certain indices or measures along with searching for patterns of relationship that

exist among the data groups. Analysis, particularly in case of survey or experimental data, involves estimating the values of unknown parameters of the population and testing of hypotheses for drawing inferences Analysis may, therefore, be categorized as:

I. Descriptive analysis: It is largely the study of distributions of one variable. This study provides us with profiles of companies, work groups, persons and other subjects on any of a multitude of characteristics such as size, composition, efficiency, preferences, etc. This sort of analysis may be in respect of one variable (described as unidimensional analysis), or in respect of two variables (described as bivariate analysis) or in respect of more than two variables (described as multivariate analysis). In this context we work out various measures that show the size and shape of distributions, along with the study of measuring relationships between two or more variables.

We may as well talk of correlation analysis and causal analysis. *Correlation analysis* studies the joint variation of two or more variables for determining the amount of correlation between two or more variables. *Causal analysis* is concerned with the study of how one or more variables affect changes in another variable. It is thus a study of functional relationships existing between two or more variables. This analysis can be termed as regression analysis. Causal analysis is considered relatively more important in experimental researches, whereas in most social and business researches our interest lies in understanding and controlling relationships between variables than with determining causes *per se* and as such we consider correlation analysis as relatively more important.

In modern times, with the availability of computer facilities, there has been a rapid development of *multivariate analysis,* which may be defined as "all statistical methods which simultaneously analyze more than two variables on a sample of observations." Usually the following analyses are involved when we make a reference of multivariate analysis:

(a) ***Multiple regression analysis:*** This analysis is adopted when the researcher has one dependent variable which is presumed to be a function of two or more independent variables. The objective of this analysis is to make a prediction about the dependent variable based on its covariance with all the concerned independent variables.

(b) ***Multiple discriminant analysis:*** This analysis is appropriate when the researcher has a single dependent variable that cannot be measured, but can be classified into two or more groups on the basis of some attribute. The object of this analysis happens to be to predict an entity's possibility of belonging to a particular group based on several predictor variables.

(c) ***Multivariate analysis of variance* (or multi-ANOVA):** This analysis is an extension of two-way ANOVA, wherein the ratio of among group variance to within group variance is worked out on a set of variables.

(d) ***Canonical analysis:*** This analysis can be used in case of both measurable and non-measurable variables for the purpose of simultaneously predicting a set of dependent variables from their joint covariance with a set of independent variables.

II. Inferential analysis: This is concerned with the various tests of significance for testing hypotheses in order to determine with what validity data can be said to indicate some conclusion or conclusions. It is also concerned with the estimation of population values. It is mainly on the basis of inferential analysis that the task of interpretation (i.e., the task of drawing inferences and conclusions) is performed.

7.9 SUMMARY

So from the above discussion it is clear that data processing mainly involves preliminary screening of the data collected, editing of data, coding of data, classification of data, tabulation of data, and analysis of data.

7.10 ANSWERS TO 'CHECK YOUR PROGRESS'

1. The steps involved in the processing of data are:

1. Editing of data, 2. Coding of data, 3. Classification of data, and 4. Tabulating of data.

2. Editing is the process of eliminating errors in the raw data collected so that the data are prepared for subsequent coding.

3. Coding is the procedure of classifying the answers to a question into meaningful categories.

4. Data collected are not directly useful for tabulation. For this the classification of data is essential. The aim of classification is to arrange data systematically for quick and easy understanding / analysis.

5. Stubs indicate the row headings or row titles. They are usually placed at the extreme left hand column.

7.11 EXERCISES AND QUESTIONS

1.What is data processing? Explain the various functions of data processing.

2.Explain the importance of data editing in marketing research.

3.What are the steps in coding data?

4.What are the bases of classifying data?

5.List out the various methods of tabulating data

6.Make a brief sketch of the various methods of analyzing data.

7.12 FURTHER READING

1. Kothari C R – Quantitative Techniques (Vikas Publishing House 3rd ed.)

2. Nargundkar R – Marketing Research Text and Cases (Tata McGraw- Hill 2002)

UNIT VIII

UNIVARIATE ANALYSIS, HYPOTHESIS TESTING AND BIVARIATE ANALYSIS

STRUCTURE

8.1 Introduction
8.2 Univariate analysis
8.3 Measures of central tendency
8.4 Measures of dispersion
8.5 Hypothesis testing
8.6 Bivariate analysis
8.7 Correlation analysis
8.8 Regression analysis
8.9 Summary
8.10 Answers to 'Check Your Progress'
8.11 Exercises and Questions
8.12 Further Reading

UNIT OBJECTIVES

The objective of this chapter is to understand:

- Methods of analysing Market Research data
- Methods of unvaried Analysis
- Hypothesis testing
- Parametric and non parametric tests
- Importance and methodology of correlation analysis.
- Importance and methodology of regression analysis.

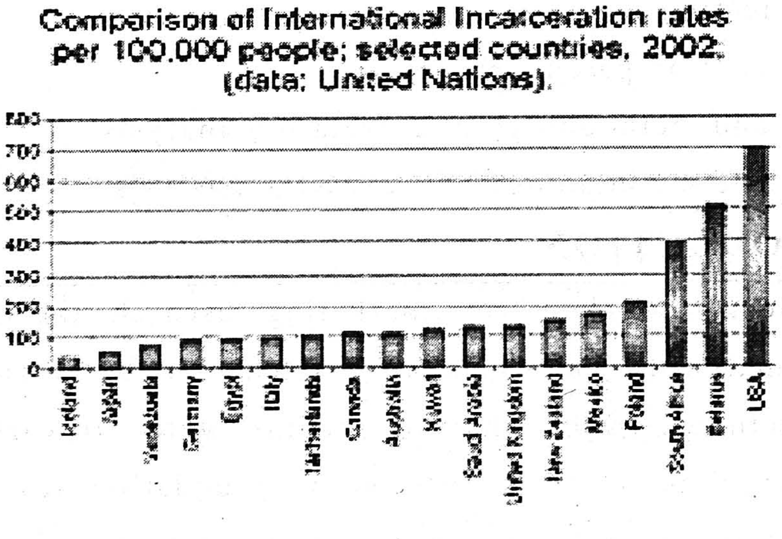
8.1 INTRODUCTION

After collection of data from marketing research, the data has to be analyzed. This can be carried out by various statistical methods. Data analysis begins with univariate analysis. Univariate analysis also is the foundation for the hypothesis testing, bivariate and multivariate analysis.

Notes

8.2 UNIVARIATE ANALYSIS

Univariate analysis is the simplest form of quantitative (statistical) analysis. The analysis is carried out with the description of a single variable and its attributes of the applicable unit of analysis. For example, if the variable age was the subject of the analysis, the researcher would look at how many subjects fall into a given age attribute categories. Univariate analysis is also used primarily for descriptive purposes. Univariate analysis is commonly used in the first stages of research, in analyzing the data at hand, before being supplemented by more advance, inferential bivariate or multivariate analysis. A basic way of presenting univariate data is to create a frequency distribution of the individual cases, which involves presenting the number of attributes of the variable studied for each case observed in the sample. This can be done in a table format, with a bar chart or a similar form of graphical representation. A sample distribution table and a bar chart for an univariate analysis are presented below (the table shows the frequency distribution for a variable "age" and the bar chart, for a variable "incarceration rate"):



Check your Progress

1) What is an univariat canalysis ?

Age range	Frequency	Percent
under 18	10	5
18–29	50	25
29–45	40	20
45–65	40	20
over 65	60	30
Valid cases: 200 Missing cases: 0		

There are several tools used in univariate analysis; their applicability depends on whether we are dealing with a continuous variable (such as age) or a discrete variable (such as gender).

8.3 MEASURES OF CENTRAL TENDENCY

In statistics, the term central tendency relates to the way in which quantitative data tend to cluster around some value. A measure of central tendency is any of a number of ways of specifying this "central value". In practical statistical analyses, the terms are often used before one has chosen even a preliminary form of analysis: thus an initial objective might be to "choose an appropriate measure of central tendency".In the simplest cases, the measure of central tendency is an average of a set of measurements, the word average being variously construed as mean, median, or other measure of location, depending on the context. However, the term is applied to multidimensional data as well as to univariate data and in situations where a transformation of the data values for some or all dimensions would usually be considered necessary: in the latter cases, the notion of a "central location" is retained in converting an "average" computed for the transformed data back to the original units. In addition, there are several different kinds of calculations for central tendency, where the

Notes

kind of calculation depends on the type of data (level of measurement).Both "central tendency" and "measure of central tendency" applies to either statistical populations or to samples from a population.

Measures of central tendency summarize the entire distribution of values as one single quantity or quality that can be thought of as the average value. Measures of central tendency are measures of the location of the middle or the center of a distribution. The mean is the most commonly used measure of central tendency. The three most commonly-used measures of central tendency are:

8.3.1 Arithmetic mean is the sum of all measurements divided by the number of observations in the data set. In mathematics and statistics, the arithmetic mean (or simply the mean) of a list of numbers is the sum of the entire list divided by the number of items in the list. If the list is a statistical population, then the mean of that population is called a population mean. If the list is a statistical sample, we call the resulting statistic a sample mean. The mean is the most commonly-used type of average and is often referred to simply as the average. The mean is the sum of the observations divided by the number of observations. The mean of a set of numbers x_1, x_2,..., x_n is typically denoted by x bar{x}. The mean is often quoted along with the standard deviation: the mean describes the central location of the data, and the standard deviation describes the spread. The arithmetic mean is commonly called the average. The sum of the values divided by the number of values--often called the "average." The mean is the sum of all the scores divided by the number of scores. where μ is the population mean and N is the number of scores. If the scores are from a sample, then the symbol M

Check your Progress

2) What is a mean?

refers to the mean and N refers to the sample size. The formula for M is the same as the formula for μ. The mean is a good measure of central tendency for roughly symmetric distributions. This can be misleading in skewed distributions since it can be greatly influenced by extreme scores. Therefore, other statistics such as the median may be more informative for distributions such as reaction time or family income that are frequently much skewed. The arithmetic mean is the "standard" average, often simply called the "mean".

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i$$

For example, the arithmetic mean of six values: 34, 27, 45, 55, 22, 34 is {34+27+45+55+22+34}\{6} = {217}\{6} =approx 36.167.

8.3.2 Median is the middle value that separates the higher half from the lower half of the data set. The median is the middle score of a distribution: half the scores are above the median and half are below the median. The median is less sensitive to extreme scores than the mean and this makes it a better measure than the mean for highly skewed distributions. The median is usually more informative than the mean.

$$M = l_1 + \frac{l_2 - l_1}{f}(m - c)$$

M = Median

l_1 = the lower limit of the class in which the median lies

l_2 = the upper limit of the class in which the median lies

f = the frequency of the class in which the median lies

m = the middle item or n/2

c = the cumulative frequency of the class preceding the one in which the median lies.

8.3.3 Mode is the most frequent value in the data set. The mode of a data sample is the element that occurs most often in the collection. For example, the mode of the sample [1, 3, 6, 6, 6, 6, 7, 7, 12, 12, 17] is 6. Given the list of data [1, 1, 2, 4, 4] the mode is not unique - the dataset may be said to be bimodal, while a set with more than two modes may be described as multimodal. The third common measure of central tendency is the mode. The advantage of the mode as a measure of central tendency is that its meaning is obvious. Further, it is the only measure of central tendency that can be used with nominal data. Some of the important characteristics of the mode are:

- It can be applied to both qualitative and quantitative distribution
- It is not affected by the extreme values in the distribution
- It can be ascertained in an open-ended distribution

Mode denotes the most frequently-occurring value (or values). The mode is the value (or values) with the highest frequency.

$$\text{Mode} = l_1 + \frac{f_1 - f_0}{(f_1 - f_0) - (f_1 - f_2)} \times i$$

l_1 = the lower value of the class in which the mode lies

f_1 = the frequency of the class in which the mode lies

f_0 = the frequency of the class preceding the modal class

f_2 = the frequency of the class succeeding the modal class

i = the class interval of the modal class

Comparison of common averages			
Type	**Description**	**Example**	**Result**
Arithmetic mean	Total sum divided by quantity of integers	(1+2+2+3+4+7+9) / 7	4
Median	Middle value that separates the greater and lesser halves of a data set	1, 2, 2, 3, 4, 7, 9	3
Mode	Most frequent number in a data set	1, 2, 2, 3, 4, 7, 9	2

8.3.4 Geometric mean is the nth root of the product of the data values. The geometric mean is an average that is useful for sets of positive numbers that are interpreted according to their product and not their sum (as is the case with the arithmetic mean) e.g. rates of growth.

$$\bar{x} = \left(\prod_{i=1}^{n} x_i\right)^{1/n}$$

For example, the geometric mean of six values: 34, 27, 45, 55, 22, 34 is:

$$(34 \cdot 27 \cdot 45 \cdot 55 \cdot 22 \cdot 34)^{1/6} = 1{,}699{,}493{,}400^{1/6} \approx 34.545.$$

The geometric mean can also be computed by:

1. taking the logarithm of each number
2. computing the arithmetic mean of the logarithms
3. raising the base used to take the logarithms to the arithmetic mean.

8.3.5 Harmonic mean - the reciprocal of the arithmetic mean of the reciprocals of the data values. The harmonic mean is an average which is useful for sets of numbers which are defined in relation to some unit, for example speed (distance per unit of time).

$$\bar{x} = n \cdot \left(\sum_{i=1}^{n} \frac{1}{x_i}\right)^{-1}$$

For example, the harmonic mean of the six values: 34, 27, 45, 55, 22, and 34 is

$$\frac{6}{\frac{1}{34} + \frac{1}{27} + \frac{1}{45} + \frac{1}{55} + \frac{1}{22} + \frac{1}{34}} = \frac{60588}{1835} \approx 33.0179836.$$

Relationship between AM, GM, and HM

AM, GM, and HM satisfy these inequalities:

$$AM \geq GM \geq HM$$

The choice of a measure of central tendency depends upon the level of measurement (nominal, ordinal, interval, or ratio) of the variable and the shape of its distribution. The mode is the only indicator of central tendency for a nominal variable. It may be computed for other types of variables as well, but is not especially useful unless there is a distinct peak, that is, when one value clearly predominates. At least an ordinal level of measurement is required for the median; the mean additionally requires an interval level of measurement. In general, the mean is preferred over the median and the median over the mode because the mean utilizes the most information about the distribution whereas the mode preserves only one piece of information. This also depends on the manner in which the variable is distributed. It is necessary to examine the entire set of values to determine which value best typifies the set as a whole. It is inadvisable to calculate a measure of central tendency for a variable without first examining its distribution.

8.4 MEASURES OF STATISTICAL DISPERSION

In statistics, statistical dispersion (also called statistical variability or variation) is variability or spread in a variable or a probability distribution. Common examples of measures of statistical dispersion are the variance, standard deviation and interquartile range. Measures of dispersion are important for describing the spread of the data, or its variation around a central value. Two distinct samples may have the same mean or median, but completely different levels of variability, or vice versa. A proper description of a set of data should include both of these characteristics. There are various methods that can be used to measure the dispersion of a dataset, each with its own set of advantages and disadvantages.

Most measures of dispersion have the same scale as the quantity being measured. In other words, if the measurements have units, such as metres or seconds, the measure of dispersion has the same units. Such measures of dispersion include:

8.4.1 Standard deviation is the standard deviation of a statistical population, a data set, or a probability distribution is the square root of its variance. Standard deviation is a widely used measure of the variability or dispersion. It shows how much variation there is from the "average" (mean) (or expected/budgeted value). It helps detect tampering of data. A low standard deviation indicates that the data points tend to be very close to the mean, whereas high standard deviation indicates that the data are spread out over a large range of values. Consider a population consisting of the following eight values:

$$2,\ 4,\ 4,\ 4,\ 5,\ 5,\ 7,\ 9.$$

The eight data points have a mean (or average) value of 5:

$$\frac{2+4+4+4+5+5+7+9}{8} = 5.$$

To calculate the population standard deviation, first compute the difference of each data point from the mean, and square the result:

$$\begin{array}{ll} (2-5)^2 = (-3)^2 = 9 & (5-5)^2 = 0^2 = 0 \\ (4-5)^2 = (-1)^2 = 1 & (5-5)^2 = 0^2 = 0 \\ (4-5)^2 = (-1)^2 = 1 & (7-5)^2 = 2^2 = 4 \\ (4-5)^2 = (-1)^2 = 1 & (9-5)^2 = 4^2 = 16 \end{array}$$

Next divide the sum of these values by the number of values and take the square root to give the standard deviation:

$$\sqrt{\frac{9+1+1+1+0+0+4+16}{8}} = 2.$$

Therefore, the above has a population standard deviation of 2.

An estimator for σ sometimes used is the standard deviation of the sample, denoted by S_n and defined as follows:

$$s_n = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2}.$$

8.4.2 Probability distribution or random variable

Let X be a random variable with mean value μ:

$$\mathrm{E}[X] = \mu.$$

Here the operator E denotes the average or expected value of X. Then the standard deviation of X is the quantity

$$\sigma = \sqrt{\mathrm{E}[(X - \mu)^2]}.$$

That is, the standard deviation σ (sigma) is the square root of the average value of (X − μ)2..In the case where X takes random values from a finite data set $x_1, x_2, \ldots, x_N$, with each value having the same probability, the standard deviation is

$$\sigma = \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + \cdots + (x_N - \mu)^2}{N}},$$

or, using summation notation,

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2}.$$

The most common estimator for σ used is an adjusted version, the sample standard deviation, denoted by "s" and defined as follows:

$$s = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2},$$

For a finite population with equal probabilities on all points, we have

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i-\overline{x})^2}=\sqrt{\frac{1}{N}\left(\sum_{i=1}^{N}x_i^2\right)-\overline{x}^2}.$$

8.4.3 Interquartile range

$$Q_3-Q_1/2$$

8.4.4 Range is the range is the length of the smallest interval which contains all the data. It is calculated by subtracting the smallest observation (sample minimum) from the greatest (sample maximum) and provides an indication of statistical dispersion. It is defined as the difference between the largest and smallest sample values. The range depends only on extreme values and provides no information about how the remaining data is distributed.

Range = L-S

Coefficient of range = $\frac{L-S}{L+S}$

8.4.5 Median absolute deviation is the median absolute deviation (MAD) is a robust measure of the variability of a univariate sample of quantitative data. It can also refer to the population parameter that is estimated by the MAD calculated from a sample. For a univariate data set $X_1, x_2, ..., X_n$, the MAD is defined as the median of the absolute deviations from the data's median

$$\text{MAD} = \text{median}_i\left(\left|X_i - \text{median}_j(X_j)\right|\right),$$

Consider the data (1, 1, 2, 2, 4, 6, 9). It has a median value of 2. The absolute deviations about 2 are (1, 1, 0, 0, 2, 4, 7) which in turn have a median value of 1 (because the sorted absolute deviations are (0, 0, 1, 1, 2, 4, 7)). So the median absolute deviation for this data is 1.

8.4.6 Average absolute deviation (or simply called average deviation). The average absolute deviation or simply average deviation of a data set is the average of the absolute

deviations and is a summary statistic of statistical dispersion or variability. It is also called the mean absolute deviation, but this is easily confused with the median absolute deviation. The average absolute deviation of a set {x1, x2, ..., xn} is

$$\frac{1}{n}\sum_{i=1}^{n}|x_i - m(X)|.$$

The choice of measure of central tendency, m(X), has a marked effect on the value of the average deviation. For example, for the data set {2, 2, 3, 4, 14}:

Measure of central tendency m(X)	Average absolute deviation
Mean = 5	$\frac{\|2-5\|+\|2-5\|+\|3-5\|+\|4-5\|+\|14-5\|}{5} = 3.6$
Median = 3	$\frac{\|2-3\|+\|2-3\|+\|3-3\|+\|4-3\|+\|14-3\|}{5} = 2.8$
Mode = 2	$\frac{\|2-2\|+\|2-2\|+\|3-2\|+\|4-2\|+\|14-2\|}{5} = 3.0$

These are frequently used (together with scale factors) as estimators of scale parameters, in which capacity they are called estimates of scale. Other measures of dispersion are dimensionless (scale-free). In other words, they have no units even if the variable itself has units. These include:

8.4.7 Coefficient of variation

In probability theory and statistics, the coefficient of variation (CV) is a normalized measure of dispersion of a probability distribution. It is defined as the ratio of the standard deviation to the mean:

Notes

$$c_v = \frac{\sigma}{\mu}$$

This is only defined for non-zero mean, and is most useful for variables that are always positive. It is also known as unitized risk or the variation coefficient. It is expressed as percentage. The coefficient of variation should only be computed for data measured on a ratio scale.

8.4.8 Quartile coefficient of dispersion

The quartile coefficient of dispersion is a descriptive statistic used to make comparisons within and between data sets. The statistic is easily computed using the first (Q_1) and third (Q_3) quartiles for each data set. Consider the following two data sets:The quartile coefficient of dispersion is

$$\frac{Q_3 - Q_1}{Q_3 + Q_1}.$$

Data set A: 2, 4, 6, 8, 10, 12, 14.

n = 7, range = 12, mean = 8, median = 8, Q1 = 4, Q3 = 12

coefficient of dispersion = 0.5

Data set B: 1.8, 2, 2.1, 2.4, 2.6, 2.9, 3

n = 7, range = 1.2, mean = 2.4, median = 2.4, Q1 = 2, Q3 = 2.9

coefficient of dispersion = 0.18

8.4.9 Variance (the square of the standard deviation)

$$\text{variance} = \sigma^2 = \frac{\Sigma(x_T - \mu)^2}{n}$$

Notes

Example: Find out the variance for the following. The table shows sco (out of 10) obtained by 20 people in a test.

Scores (x)	Frequency (f)
1	0
2	1
3	1
4	3
5	2
6	5
7	5
8	2
9	0
10	1

Solution:

Scores (x)	Frequency (f)	fx	fx^2
1	0	0	0
2	1	2	4
3	1	3	9
4	3	12	48
5	2	10	50
6	5	30	180
7	5	35	245
8	2	16	128
9	0	0	0
10	1	10	100
	20	118	764

Notes

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f}$$

$$\text{Variance}, \sigma^2 = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2$$

$$\text{Variance} = \frac{764}{20} - \left(\frac{118}{20}\right)^2$$

$= 38.2 - 34.81 = \mathbf{3.39}$

8.5 HYPOTHESIS TESTING

A statistical hypothesis is an assumption about a population parameter. This assumption may or may not be true. The best way determine whether a statistical hypothesis is true is to examine the entire population. Since this is often impractical, researchers typically examine a random sample from the population. If the sample data are consistent with the statistical hypothesis, the hypothesis is accepted; if not, the hypothesis is rejected. There are two types of statistical hypothesis.

Null hypothesis- The null hypothesis is usually the hypothesis that ample observations result purely from chance effects.

Alternative hypothesis- The alternative hypothesis is the hypothesis that sample observations are influenced by some non-random cause. For example, suppose we wanted to determine whether a coin was fair and balanced. A null hypothesis might be that half the flips would result in Heads and half, in Tails. The alternative hypothesis might be that the number of Heads and Tails would be very different. Suppose we flipped the coin 50 times, resulting in 40 Heads and 10 Tails. Given this result, we would be inclined to reject the null hypothesis and accept the alternative hypothesis. The term "null" means nothing or invalid. It may be written as:

$$\mathbf{H_0: \mu = \mu_0}$$

Where, H_0 is the null hypothesis and μ_0 is the mean of the population. The alternative hypothesis is

Check your Progress

2) What is H_0?

H_a: μ not equal to μ_0

The rejection of the null hypothesis will show that the mean of the population is not μ_0. If this happens, then it implies that the alternative hypothesis is accepted. There can be more than two or more alternative hypothesis though only one can be tested at a time against the null hypothesis.

8.5.1 Hypothesis testing steps

Statisticians follow a formal process to determine whether to accept or reject a null hypothesis, based on sample data. This process, called hypothesis testing, consists of four steps.

- **Formulating the hypothesis**: The first step in hypothesis testing is to formulate the hypothesis to be tested. This means stating the null hypothesis and the alternative hypothesis.
- **Identifying the test statistic**: The test statistic is a statistic that will be used by the researcher to determine whether the null hypothesis should be accepted or rejected. Typically, the test statistic is the sample estimate of the population parameter in the null hypothesis. Therefore, since we are testing a hypothesis about a population mean, the test statistic will be the sample mean. When the hypothesis pertains to a large sample (30 or more), the z-test implying normal distribution is used. When a sample is small (less than 30), the use of the z-test will be inappropriate. Instead the t-test will be more suitable. The test criteria frequently used in hypothesis testing are Z, t, F and chi square analysis
- **Formulating a decision rule**: The decision rule consists of two parts: (1) a test statistic and (2) a range of values, called the region of acceptance. The decision rule determines whether a null hypothesis is accepted or rejected. If the test statistic falls within the region of acceptance, the null hypothesis is accepted; otherwise, it is rejected.
- **Accepting or rejecting the null hypothesis**: Once the region of acceptance is defined, the null hypothesis can be tested against sample data. The test statistic is computed. For example consider that the test

Notes

statistic is the sample mean. If the sample mean falls within the region of acceptance, the null hypothesis is accepted; if not, it is rejected.

8.5.2 Decision Rules

There are four possibilities that can arise when a hypothesis is tested:

1. The hypothesis is true but our test leads to its rejection.
2. The hypothesis is false but our test leads to its acceptance.
3. The hypothesis is true and our test leads to its acceptance.
4. The hypothesis is false and our test leads to its rejection. Out of these four, the first two lead to an error in decision. The first possibility leads to a Type I error and the second possibility leads to a Type II error. This can be shown as follows:

STATE OF NATURE

DECISION		H_0 True	H_0 False
	Retain H_0	Correct Retention	Type II Error
	Reject H_0	Type I Error	Correct Rejection

The decision rule is a procedure that a researcher uses to decide whether to accept or reject the null hypothesis. There are two types of errors that can result from a decision rule.

- **Type I error**. A Type I error occurs when the researcher rejects a null hypothesis when it is true. The probability of committing a Type I error is called the **significance level**. This probability is also called **alpha**, and is often denoted by á.
- **Type II error**. A Type II error occurs when the researcher accepts a null hypothesis that is false. The probability of committing a Type II error is called **Beta**, and is often denoted by â. The probability of *not* committing a Type II error is called the **Power** of the test. In practice, the decision rule has two parts: (1) a test statistic and (2) a range of values. The range of values is called the **region of acceptance**. The region of acceptance is defined so that the chance of making a Type I error is equal to the significance level. If the test statistic falls within the region of

acceptance, the null hypothesis is accepted. Note: The set of values outside the region of acceptance is called the **region of rejection**. If the test statistic falls within the region of rejection, the null hypothesis is rejected. In such cases, we say that the hypothesis has been rejected at the á level of significance.

8.5.3 One-Tailed and Two-Tailed Tests

To understand the difference between these look at the table below. It shows three sets of hypothesis. Each makes a statement about how the populations mean μ is related to a specified value *M*.

Set	Null hypothesis	Alternative hypothesis	Number of tails
1	μ = M	μ not equal to M	2
2	μ > M	μ < M	1
3	μ < M	μ > M	1

The first set of hypotheses (Set 1) is an example of a two-tailed test, since an extreme value on either side of the sampling distribution would cause a researcher to reject the null hypothesis. The other two sets of hypotheses Sets 2 and 3) are one-tailed tests, since an extreme value on only one side of the sampling distribution would cause a researcher to reject the null hypothesis. A test of a statistical hypothesis, where the region of rejection is on only one side of the sampling distribution, is called a **one-tailed test**. For example, suppose the null hypothesis states that the mean is less than or equal to 10. The alternative hypothesis would be that the mean is greater than 10. The region of rejection would consist of a range of numbers located on the right side of sampling distribution; that is, a set of numbers greater than 10. A test of a statistical hypothesis, where the region of rejection is on both sides of the sampling distribution, is called a **two-tailed test**. For example, suppose the null hypothesis states that the mean is equal

to 10. The alternative hypothesis would be that the mean is less than 10 or greater than 10. The region of rejection would consist of a range of numbers located on both sides of sampling distribution; that is, the region of rejection would consist partly of numbers that were less than 10 and partly of numbers that were greater than 10.

8.5.4 Illustration for Normal distribution test (Z test)

Intelligence test given to two groups of boys and girls gave the following information:

	Mean Score	S.D.	Number
Girls	75	10	50
Boys	70	12	100

Is the difference in the mean scores of boys and girls statistically significant?

Solution: Let us take the hypothesis that the difference in the mean score of boys and girls is not significant, i.e., $\mu_1 = \mu_2$.

We are given

$\overline{x_1} = 75, \overline{x_2} = 70, s_1^2 = 100, s_2^2 = 144,\ n_1 = 50, n_2 = 100.$

The appropriate statistic to be used here is given by

$$z = \frac{(\overline{x_1} - \overline{x_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Since $\sigma_1^2 = s_1^2$; $\sigma_2^2 = s_2^2$ and $\mu_1 = \mu_2$

$$z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$= \frac{75 - 70}{\sqrt{\frac{100}{50} + \frac{144}{100}}} = \frac{5}{\sqrt{3.44}} = \frac{5}{1.855} = 2.695$$

Since the computed value z=2.695 is greater than the critical value of z=2.58 at 1% level of significance, therefore,

the hypothesis is rejected. Hence the difference in the mean score of boys and girls is statistically significant.

8.5.5 Illustration for t – test

Two salesmen A and B are working in a certain district. From a sample survey conducted by the Head Office, the following results were obtained. State whether there is any significant difference in the average sales between the two salesmen.

	A	B
No. of Sales	20	18
Average Sales (in Rs)	170	205
Standard Deviation (in Rs.)	20	25

Solution: Null hypothesis $H_0 : \mu_1 = \mu_2$ i.e., there is no difference in the average sales between the two salesmen. Applying t-test

$$t = \frac{\overline{x_1} - \overline{x_2}}{s}\sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

$$s = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

$$\overline{x_1} = 170, \overline{x_2} = 205, n_1 = 20, n_2 = 18, s_1 = 20, s_2 = 25$$

$$= \sqrt{\frac{19(20)^2 + (17)(25)^2}{20 + 18 - 2}} = \sqrt{\frac{7600 + 10625}{36}}$$

$$= \sqrt{\frac{18225}{36}} = \sqrt{506.25} = 22.5$$

$$t = \frac{170 - 205}{22.5}\sqrt{\frac{20 * 18}{20 + 18}} = \frac{-35}{22.5}\sqrt{\frac{360}{38}} = -1.56 * 3.08 = -4.8$$

The table value of t at 5% level of significance for 36d.f. (when d.f. is more than 30, than t distribution is the same as

normal distribution) is 1.9. Since the computed value of t is much more than the table value, we reject the null hypothesis. Thus we conclude that there is significance difference in the average sales between the two salesmen.

8.5.5 Illustration for F- Test

Two random samples drawn from normal populations are:

Sample 1: 20 16 26 27 23 22 18 24 25 19

Sample 2: 27 33 42 35 32 34 38 28 41 43 30 37

Obtain estimates of the variance of the populations have the same variance.

Solution: Let us take the null hypothesis that two population have the same variance. Applying F-test

Sample I			Sample II		
x_1	$(x_1 - \bar{x}_1)$	$(x_1 - \bar{x}_1)^2$	x_2	$(x_2 - \bar{x}_2)$	$(x_2 - \bar{x}_2)^2$
20	-2	4	27	-8	64
16	-6	36	33	-2	4
26	+4	16	42	+7	49
27	+5	25	35	0	0
23	+1	1	32	-3	9
22	0	0	34	-1	1
18	-4	16	38	+3	9
24	+2	4	28	-7	49
25	+3	9	41	+6	36
19	-3	9	43	+8	64
			30	-5	25
			37	+2	4
$\Sigma x_1 = 220$		$\Sigma (x_1 - \bar{x}_1)^2 = 120$	$\Sigma x_2 = 420$		$\Sigma (x_2 - \bar{x}_2)^2 = 314$

$$s_1^2 = \frac{\Sigma (x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{120}{9} = 13.333;$$

$$s_2^2 = \frac{\Sigma (x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{314}{11} = 28.545$$

$$F = \frac{s_1^2}{s_2^2} = \frac{13.333}{28.545} = 0.467$$

Since numerator is greater than denominator, therefore,

$$F = \frac{28.545}{13.33} = 2.14$$

The critical value of F for $V_1=9$ and $V_2=11$ at 5% level is 4.63. Since the calculated value of F is less the table value, therefore, there is no reason to reject the null hypothesis. Hence, it may be concluded that the two populations have the same variance.

8.5.6 Illustration for χ^2 – test

A sample analysis of examination results of 200 MBA's vas made. It was found that 46 students have failed, 68 secured a third division, 62 secured a second division and rest were placed in the first division. Are these figures commensurate with the general examination result which is in the ratio of 2:3:3:2, for various categories respectively? (Given for $v = 3$ $\chi^2_{0.05} = 7.81$)

Solution:

Category	O	E	$(O-E)^2$	$(O-E)^2/E$
Failed	46	40	36	0.900
Third Division	68	60	64	1.067
Second Division	62	60	4	0.067
First Division	24	40	256	6.400
				$\Sigma[(O-E)^2/E]=8.434$

Let us take the hypothesis that there is no difference in the observed and expected results. On the basis of ratio 2:3:3:2, the expected number of students failing, getting third division, second division, and first division, should be 200x2/10 = 40, 60, 60,40 respectively. Applying χ2test: The table value of χ^2 for 3 d.f. at 5% level of significance is 7.81. The calculated value of χ^2 is greater than the table value. The hypothesis does not hold true. Hence the given results are not commensurate with the general examination results.

8.5.7 Parametric and Non-Parametric Tests

Parametric statistics is a branch of statistics that assumes data come from a type of probability distribution and makes inferences about the parameters of the distribution. Most well-known elementary statistical methods are parametric. Generally speaking parametric methods make more assumptions than non-parametric methods. If those extra assumptions are correct, parametric methods can produce more accurate and precise estimates. They are said to have more statistical power. However, if those assumptions are incorrect, parametric methods can be very misleading. For that reason they are often not considered robust. On the other hand, parametric formulae are often simpler to write down and faster to compute. In some, but definitely not all, cases their simplicity makes up for their non-robustness, especially if care is taken to examine diagnostic statistics. Because parametric statistics require a probability distribution, they are not distribution-free.

Non-parametric statistics is a branch of statistics concerned with non-parametric statistical models and non-parametric statistical tests. Non-parametric statistics are statistics that do not estimate population parameters. In contrast, see parametric statistics. Non-parametric models differ from parametric models in that the model structure is not specified a priori but is instead determined from data. The term nonparametric is not meant to imply that such models completely lack parameters but that the number and nature of the parameters are flexible and not fixed in advance. Nonparametric models are therefore also called distribution free. Non-parametric (or distribution-free) inferential statistical methods are mathematical procedures for statistical hypothesis testing which, unlike parametric statistics, make no

Notes

assumptions about the frequency distributions of the variables being assessed. Nonparametric tests have less power than the appropriate parametric tests, but are more robust when the assumptions underlying the parametric test are not satisfied.

8.6 BIVARIATE ANALYSIS

In bivariate analysis, the hypothesis of "association" and causality are tested. In its simplest form, association simply refers to the extent to which it becomes easier to know/predict a value for the Dependent Variable (DV) if we know a case's value on the Independent Variable (IV). This association could be understood by a measure of association. A measure of association often ranges between –1 and 1. Where the sign of the integer represents the "direction" of correlation (negative or positive relationships) and the distance away from 0 represents the degree or extent of correlation – the farther the number away from 0, the higher or "more perfect" the relationship is between the IV and DV.Measures of association and statistical significance that are used vary by the level of measurement of the variables analyzed.

8.7 CORRELATION ANALYSIS

The Statistical tool with the help of which these relationships between two or more than two variables is studied is called Correlation. The measure of correlation called the coefficient of correlation (denoted by the symbol r) summarizes in one figure the direction used in measuring the closeness of the relationship between the variables. The problem of analyzing the relation between different series should be broken down into three steps:

Check your Progress

2) What is correla-tion?

1. Determining whether a relation exists and, if it does, measuring it;
2. Testing whether it is significant; and

3. Establishing the cause-and-effect relation, if any.

8.7.1 Types of Correlation

Correlation is described or classified in several different ways. Three of the most important are;

1. Positive and negative;
2. Simple, partial and multiple; and
3. Linear and non-linear.

8.7.2 Methods of Studying Correlation

The following are the important methods of ascertaining whether two variables are correlated or not:

1. Scatter Diagram method
2. Karl Pearson's Coefficient of Correlation
3. Spearman's Rank Correlation Coefficient and
4. Method of Least Squares.

8.7.3 Karl Pearson's Coefficient of Correlation

Deviation from mean formula

$$r = \frac{\Sigma(X-\overline{X})(Y-\overline{Y})}{\sqrt{\Sigma(X-\overline{X})^2}\sqrt{\Sigma(Y-\overline{Y})^2}}$$

Formula for taking values as it is

$$r = \frac{N\sum XY - \sum X \sum Y}{\sqrt{N\sum X^2 - (\Sigma X)^2}\sqrt{N\sum Y^2 - (\Sigma Y)^2}}$$

Using assumed mean method formula

$$r = \frac{N\sum d_x d_y - \sum d_x d_y}{\sqrt{N\sum d_x^2 - (\sum d_x)^2}\sqrt{N\sum d_y^2 - (\sum d_y)^2}}$$

Check your Progress

5) What is acatter diagram?

8.7.4 Spearman Rank Correlation Coefficient

$$R = 1 - \frac{6\sum D^2}{N^3 - N}$$

Where, $D = R_1 - R_2$ = Difference between the ranks for each pair of observations

R = Coefficient of rank correlation

N = Number of paired observations

The value of the rank correlation coefficient lies between +1 and -1. When R is +1, there is complete agreement in the order of the ranks and the ranks are in the same direction. When R is -1, there is complete agreement in the order of the ranks and they are in the opposite directions.

8.7.5 Rank correlation for ranks repeated

$$R = 1 - \frac{\left\{\sum D^2 + \frac{1}{12}(m_1^3 - m_1) + \frac{1}{12}(m_2^3 - m_2) + \ldots\right\}}{N^3 - N}$$

8.8 REGRESSION ANALYSIS

Regression analysis is a statistical device with the help of which we are in a position to estimate or predict the unknown values of one variable from known values of another variable.

One of the most frequently used techniques in economics and business research, to find a relation between two or more variables that are related causally, is regression analysis". - *Taro Yamane.*

"Regression analysis attempts to establish the 'nature of the relationship' between variables – that is, to study the functional relationship between the variables and thereby provide a mechanism for prediction, or forecasting. - *Ya-Lunchou.*

"The term regression analysis refers to the methods by which estimates are made of the values of a variable from a knowledge

of the value of one or more other variables and to the measurement of the errors involved in this estimation process"*Morris Hamburg.*

Uses of regression analysis:

- It is useful to estimate the relationship between two variables.
- It is useful for prediction of unknown value.
- It is widely used in social sciences like economics, natural and physical sciences.
- It is useful to forecast the business situations.
- It is useful to estimate the error in sampling.
- It is useful to calculate correlation co-efficient and co-efficient of determination.

8.8.1 Difference between Correlation and Regression:

Correlation	**Regression**
Relationship between two or more variable.	It is a mathematical measure showing the average relationship between variables.
X and Y are random variable	X is a random variable and Y is a fixed variable
It gives limited information after verifying the relationship between the variables.	It is used for prediction on one value, in relationship to the other given value.
The range of relationship lies between -1 and +1.	Regression value is an absolute figure.
It studies linear relationship between the variables.	It studies linear and non-linear relationship between the variables
If the coefficient of correlation is positive, then the two variable are positively correlated and vice versa.	The regression coefficient explains that the decrease in one variable is associated with the increase in the other variable.

Notes

8.8.2 Regression lines:

If we take two variables X and Y we have two regression lines:

- Regression of X on Y
- Regression of Y on X

Regression equation of Y on X

$$Y = a + bX$$

To determine the values of a and b, we shall solve the normal equations:

$$\sum Y = Na + b\sum X$$

$$\sum XY = a\sum X + b\sum X^2$$

Regression equation of X on Y

$$X = a + bY$$

To determine the values of a and b, we shall solve the normal equations:

$$\sum X = Na + b\sum Y$$

$$\sum XY = a\sum Y + b\sum Y^2$$

8.8.3 Deviations taken from Arithmetic Means of X and Y

Regression equation of Y on X:

$$Y - \overline{Y} = b_{yx}(X - \overline{X})$$

$$b_{yx} = \frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sum(X-\overline{X})^2}$$

$$\text{Regression coefficient of y on x} = b_{yx} = \frac{N\sum d_x d_y - \sum d_x \sum d_y}{N\sum d_x^2 - (\sum d_x)^2}$$

Regression equation of X on Y:

$$(X - \overline{X}) = b_{xy}(Y - \overline{Y})$$

$$\text{Regression coefficient of x on y} = b_{xy} = \frac{N\sum d_x d_y - \sum d_x \sum d_y}{N\sum d_y^2 - (\sum d_y)^2}$$

Notes

8.9 SUMMARY

This unit dealt with the measures of central tendency which are the arithmetic mean, the median and the mode. This was followed by the measures of dispersion; the standard deviation and the coefficient of variation. The measures of central tendency are used to estimate "normal" values of a dataset. Measures of dispersion are important for describing the spread of the data, or its variation around a central value. Two distinct samples may have the same mean or median, but completely different levels of variability, or vice versa.

A statistical hypothesis is an assumption about a population parameter. This assumption may or may not be true. There are two types of statistical hypothesis.

1. Null hypothesis.
2. Alternative hypothesis.

This process, called hypothesis testing, consists of four steps: Formulating the hypothesis, identifying the test statistic, formulating a decision rule and accepting or rejecting the null hypothesis.

In bivariate analysis, the hypothesis of "association" and causality are tested. In its simplest form, association simply refers to the extent to which it becomes easier to know/predict a value for the Dependent Variable (DV) if we know a case's value on the Independent Variable (IV). Correlation is a statistical technique used for measuring the relationship or interdependence of two or more variables. Regression analysis refers to the technique for deriving an equation that relates the dependent variable to one or more variables. It is used to predict one variable on the basis of another and hence helps in bringing out the causal relationship between two or more variables.

Notes

8.10 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Univariate analysis is carried out with the description of a single variable and its attributes of the applicable unit of analysis.

2. The mean is the sum of the observations divided by the number of observations.

3. H_0 is the null hypothesis, it is generally stated as follows- there no difference between population parameter and sample statistic..

4. The relationships between two or more than two variables are called Correlation.

5. The Scatter Diagram is a tool for determining the potential correlation between two different sets of variables, i.e., how one variable changes with the other variable. This diagram simply plots pairs of corresponding data from two Variables.

8.11 EXERCISES AND QUESTIONS

1. What are the measures of central tendency?
2. How you measure variance?
3. How are the measures of central tendency different from the measures of dispersion?
4. What are the steps involved in testing a hypothesis?
5. What are Type I error and Type II errors?
6. What are one-tailed and two-tailed tests?
7. What is correlation analysis?
8. What is regression analysis?

8.12 FURTHER READINGS

1. Paul. E. Green, Donald. S. Tull, Gerald Albaum, *Research for Marketing Decisions*, Prentice Hall of India Pvt. Ltd. New Delhi.

2. David. A. Aaker, V. Kumar, George.S.Day, *Marketing Research*, John Wiley & Sons Inc, Singapore.

UNIT IX

MULTIVARIATE ANALYSIS

STRUCTURE

9.1 Introduction
9.2 Various multivariate analysis methods
9.3 Correlation Analysis
9.4 Regression Analysis
9.5 Factor Analysis
9.6 Cluster Analysis
9.7 Correspondence Analysis
9.8 Conjoint Analysis
9.9 CHAID Analysis
9.10 Discriminant Analysis
9.11 Multidimensional Scaling
9.12 Structural Equation Modeling
9.13 Canonical Correlation
9.14 Analysis of Variance
9.15 Logistic Regression Analysis
9.16 Summary
9.17 Answers to 'Check Your Progress'
9.18 Exercises and Questions
9.19 Further Readings

9.1 UNIT OBJECTIVE

The objective of this unit is:

- To understand Multivariate analysis
- To study about analysis of variation and Multiple Regression Analysis
- To understand Discriminant Analysis, Conjoint Analysis
- To study about Factor Analysis
- To know about Cluster Analysis and Multi-dimensional Scaling.

NOTES

9.2 INTRODUCTION

Multivariate analysis is the analysis of the simultaneous relationships among three or more phenomena. In Univariate analysis the focus is on the level (average) and distribution (variance) of the phenomenon, in a bivariate analysis the focus shifts to the degree of relationships (correlations or co variances) between the phenomena. In a multivariate analysis, the focus shifts from paired relationships to more complex simultaneous relationships among phenomena. The purpose of this chapter is to provide an executive understanding of 13 multivariate analysis techniques, resulting in an understanding of the appropriate uses for each of the techniques. This is not a discussion of the underlying statistics of each technique; it is a field guide to understanding the types of research questions that can be formulated and the capabilities and limitations of each technique in answering those questions. In order to understand multivariate analysis, it is important to understand some of the terminology. A variate is a weighted combination of variables. The purpose of the analysis is to find the best combination of weights. Nonmetric data refers to data that are either qualitative or categorical in nature. Metric data refers to data that are quantitative, and interval or ratio in nature.

9.3 VARIOUS MULTIVARIATE ANALYSIS METHODS

Before launching into an analysis technique, it is important to have a clear understanding of the form and quality of the data. The form of the data refers to whether the data are nonmetric or metric. The quality of the data refers to how normally distributed the data are. The first few techniques discussed are sensitive to the linearity, normality, and equal variance assumptions of the data. Examinations of distribution, skewness, and kurtosis are helpful in examining distribution. Also, it is important to understand the magnitude of missing values in observations and to determine whether to ignore them or impute values to the missing observations. Another data quality measure is outliers, and it is important to determine whether the outliers should be removed. If they are kept, they may cause a distortion to the data; if they are

Check your Progress

1) What is a multivariate analysis?

eliminated, they may help with the assumptions of normality. The key is to attempt to understand what the outliers represent. The important multivariate tools are:

1. Correlation Analysis
2. Regression Analysis
3. Factor Analysis
4. Cluster Analysis
5. Correspondence Analysis
6. Conjoint Analysis
7. CHAID Analysis
8. Discriminant Analysis
9. Multidimensional Scaling
10. Structural Equation Modeling
11. Canonical Correlation
12. Analysis of Variance
13. Logistic Regression Analysis

9.3 CORRELATION ANALYSIS

Correlation analysis, expressed by correlation coefficients, measures the degree of linear relationship between two variables. While in regression the emphasis is on predicting one variable from the other, in correlation the emphasis is on the degree to which a linear model may describe the relationship between two variables. The correlation coefficient may take on any value between + and - 1. The sign of the correlation coefficient (+, -) defines the direction of the relationship, either positive or negative. A positive correlation coefficient means that as the value of one variable increases, the value of the other variable increases; as one decreases the other decreases. A negative correlation coefficient indicates that as one variable increases, the other decreases, and vice-versa. The absolute value of the correlation coefficient measures the

NOTES

strength of the relationship. A correlation coefficient of r=0 .50 indicates a stronger degree of linear relationship than one of r=0.40. Thus a correlation coefficient of zero (r=0.0) indicates the absence of a linear relationship and correlation coefficients of r=+1.0 and r=-1.0 indicate a perfect linear relationship. The scatter plots presented below perhaps best illustrate how the correlation coefficient changes as the linear relationship between the two variables is altered. When r=0.0 the points scatter widely about the plot, the majority fall roughly in the shape of a circle. As the linear relationship increases, the circle becomes more and more elliptical in shape until the limiting case is reached (r=1.00 or r=-1.00) and all the points fall on a straight line.

9.4 MULTIPLE REGRESSION ANALYSIS

Multiple regressions is the most commonly utilized multivariate technique. It examines the relationship between a single metric dependent variable and two or more metric independent variables. The technique relies upon determining the linear relationship with the lowest sum of squared variances; therefore, assumptions of normality, linearity, and equal variance are carefully observed. The beta coefficients (weights) are the marginal impacts of each variable, and the size of the weight can be interpreted directly. Multiple regression is used to account for (predict) the variance in an interval dependent, based on linear combinations of interval, dichotomous, or dummy independent variables. Multiple regression can establish that a set of independent variables explains a proportion of the variance in a dependent variable at a significant level (through a significance test of R^2), and can establish the relative predictive importance of the independent variables. Multiple regression is often used as a forecasting tool. The multiple regression equation takes the form

Check your Progress

2) What does R^2 denote?

$$y = a + b_1x_1 + b_2x_2 + ... + b_nx_n$$

where y is the dependent variable which is to be predicted, x_1, x_2 and x_n are the n known variables on which the predictions are to be based and a, b_1, b_2, ...b_n are parameters, the values of which are to be determined by the

NOTES

methods of least squares. With two independent variables the prediction of Y is expressed by the following equation:

$$Y = a + b_1X_1 + b_2X_2$$

where a is the trend value at the time origin, b_1 is the slope of the independent variable X_1, and b_2 is the slope of the independent variable X_2. The value of a, b_1 and b_2 can be determined by solving the following three normal equations simultaneously.

$$\Sigma Y = na + b_1 \Sigma X_1 + b_2 \Sigma X_2 \quad \cdots\cdots (1)$$

$$\Sigma X_1Y = a\Sigma X_1 + b_1 \Sigma X^2_1 + b_2 \Sigma X_1X_2 \quad \cdots\cdots (2)$$

$$\Sigma X_2Y = a\Sigma X_2 + b_1 \Sigma X_1X_2 + b_2 \Sigma X^2_2 \quad \cdots\cdots (3)$$

$$r^2 = \frac{Ó(Y_i - \overline{Y})^2 - (Y_i - Y_c)^2}{(Y_i - \overline{Y})^2}$$

Regression analysis measures the strength of a relationship between a variable you try to explain (e.g. overall customer satisfaction) and one or more explaining variables (e.g. satisfaction with product quality and price). While correlation provides a single numeric summary of a relation (called the correlation coefficient), regression analysis results in a "prediction" equation. The equation describes the relation between the variables. If the relationship is strong (expressed by the Rsquare value), it can be used to predict values of one variable given the other variables have known values. E.g. how will the overall satisfaction score change if satisfaction with product quality goes up from 6 to 7? Regression analysis is typically used for

(i) Customer Satisfaction & Employee Satisfaction studies to answer questions such as "which product dimensions contribute most to someone's overall satisfaction or loyalty to the brand". This is often referred to as Key Drivers Analysis.

(ii) To simulate the outcome when actions are taken. e.g. what will happen to the satisfaction score when product availability is improved?

Check your Progress

3) Define regression

9.5 FACTOR ANALYSIS

Factor analysis is a name given to a class of techniques whose purpose is data reduction and summarization. The data from market research are vast and factor analysis helps in reducing the number of variables. Factor analysis is an explorative technique. Factor analysis was invented nearly 100 years ago by psychologist Charles Spearman, who hypothesized that the enormous variety of tests of mental ability--measures of mathematical skill, vocabulary, other verbal skills, artistic skills, logical reasoning ability, etc.--could all be explained by one underlying "factor" of general intelligence that he called g. He hypothesized that if g could be measured and you could select a subpopulation of people with the same score on g, in that subpopulation you would find no correlations among any tests of mental ability. In other words, he hypothesized that g was the only factor common to all those measures. The objectives of factor analysis are simplifying the data by reducing a large number of variables to a set of a small number of variables and analyzing the interdependence of relationship among a total set of variables. Factor analysis can be used in several ways as given below:

1. It brings out the hidden dimensions relevant to a researcher among product preferences.
2. Helps to find out relationships among observed values.
3. Used when the data is large and has to be simplified and condensed.

The limitations of factor analysis are:

- It is a complicated tool and should be used if the researcher has a good understanding of the technique.
- The reliability of the results is some times questionable.
- It suitability depends on the judgment of the researcher.

Factor analysis is used in the case of exploratory research and has to be used were the concepts are well formulated and tested. When there are many variables in a research design, it is often helpful to reduce the

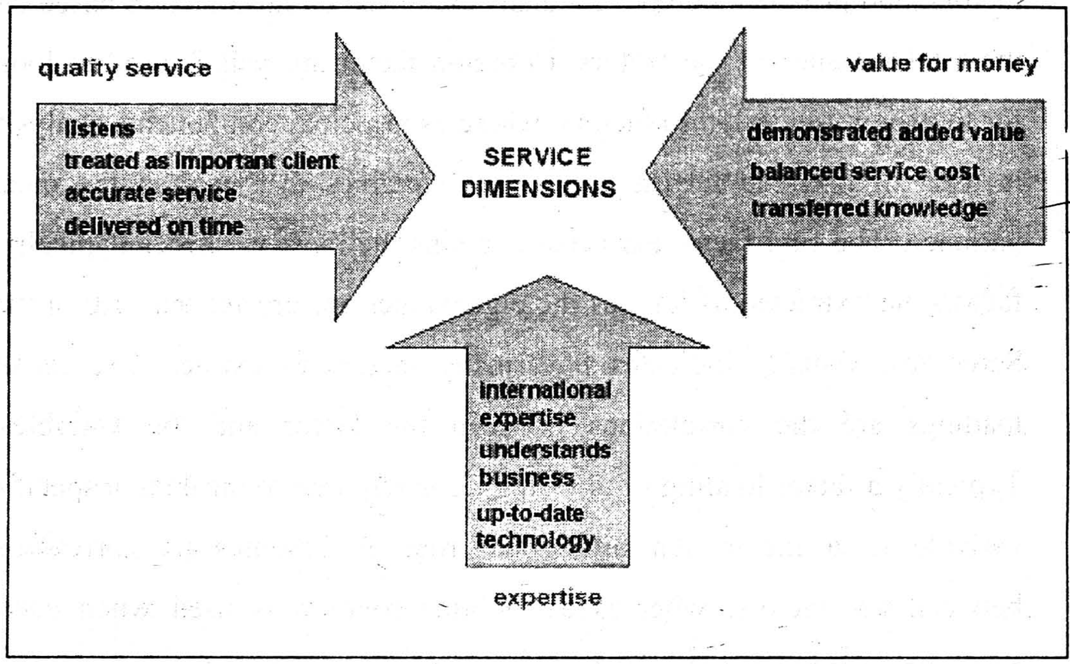
variables to a smaller set of factors. This is an independence technique, in which there is no dependent variable. Rather, the researcher is looking for the underlying structure of the data matrix. Ideally, the independent variables are normal and continuous, with at least 3 to 5 variables loading onto a factor. The sample size should be over 50 observations, with over 5 observations per variable. Multicollinearity is generally preferred between the variables, as the correlations are key to data reduction. Kaiser's Measure of Statistical Adequacy (MSA) is a measure of the degree to which every variable can be predicted by all other variables. An overall MSA of .80 or higher is very good, with a measure of under .50 deemed poor. There are two main factor analysis methods: common factor analysis, which extracts factors based on the variance shared by the factors, and principal component analysis, which extracts factors based on the total variance of the factors. Common factor analysis is used to look for the latent (underlying) factors, where as principal components analysis is used to find the fewest number of variables that explain the most variance. The first factor extracted explains the most variance. Typically, factors are extracted as long as the eigenvalues are greater than 1.0 or the Scree test visually indicates how many factors to extract. The factor loadings are the correlations between the factor and the variables. Typically a factor loading of .4 or higher is required to attribute a specific variable to a factor. An orthogonal rotation assumes no correlation between the factors, whereas an oblique rotation is used when some relationship is believed to exist.

Factor analysis aims to describe a large number of variables or questions by only using a reduced set of underlying variables, called factors. It explains a pattern of similarity between observed variables. Questions which belong to one factor are highly correlated with each other. Unlike cluster analysis, which classifies respondents, factor analysis groups variables. There are two types of factor analysis: exploratory and confirmatory. Exploratory factor analysis is driven by the data, i.e. the

Check your Progress

4) What is factor analysis?

data determines the factors. Confirmatory factor analysis, used in structural equation modelling, tests and confirms hypotheses. Factor analysis is often used in customer satisfaction studies to identify underlying service dimensions, and in profiling studies to determine core attitudes. For example, as part of a national survey on political opinions, respondents may answer three separate questions regarding environmental policy, reflecting issues at the local, regional and national level. Factor analysis can be used to establish whether the three measures do, in fact, measure the same thing. It is can also prove to be useful when a lengthy questionnaire needs to be shortened, but still retain key questions. Factor analysis will indicate which questions can be omitted without losing too much information.



9.6 CLUSTER ANALYSIS

Cluster analysis is a technique used to segment a market. It is used to classify a person or object into a small number or mutually exclusive and exhaustive groups. Its object is to sort cases (people, things, events, etc) into groups, or clusters, so that the degree of association is strong between members of the same cluster and weak between members of different clusters. Each cluster thus describes, in terms of the data collected, the class to which its members belong; and this description may

be abstracted through use from the particular to the general class or type. CA lacks an underlying body of statistical theory and is heuristic in nature. Cluster analysis requires decisions to be made by the user relating to the calculation of clusters, decisions which have a strong influence on the results of the classification. CA is useful to classify groups or objects and is more objective than subjective. Cluster analysis, like factor analysis and multi dimensional scaling, is an interdependence technique: it makes no distinction between dependent and independent variables. The entire set of interdependent relationships is examined. It is similar to multi dimensional scaling in that both examine interobject similarity by examining the complete set of interdependent relationships. The difference is that multi dimensional scaling identifies underlying dimensions, while cluster analysis identifies clusters. Cluster analysis is the obverse of factor analysis. Whereas factor analysis reduces the number of variables by grouping them into a smaller set of factors, cluster analysis reduces the number of observations or cases by grouping them into a smaller set of clusters. In marketing, cluster analysis is used for:

- Segmenting the market and determining target markets
- Product positioning and New Product Development
- Selecting test markets

Example: A supermarket might gather data on all of their existing customers and survey them regarding their buying criteria relative to the product line. They could then use cluster analysis to group customers with similar buying patterns together. This type of cluster analysis, also known as market segmentation, is performed at increasing rates, due to the advent of high-speed computers and the ready availability of demographic data. Based on the broader descriptions of individuals within each cluster, the retail managers could make decisions that would be appropriate for the individuals within. Clusters for this example might include:

- Price-sensitive shoppers
- Indifferent shoppers

- Quality-focused shoppers
- High-end status shoppers
- Monthly shopper

Clustering methods may be top down and employ logical division, or bottom up and undertake aggregation. Aggregation procedures which are based upon combining cases through assessment of similarities are the most common and popular will be the focus of this section. Care should be taken that groups (classes) are meaningful in some fashion and are not arbitrary or artificial. To do so the clustering techniques attempt to have minimal internal variation as compared to maximal variation between groups. Homogeneous and distinct groups are delineated based upon assessment of distances or an F-test.

9.6.1 Steps in Cluster Analysis:

The two key steps within cluster analysis are the measurement of distances between objects and to group the objects based upon the resultant distances (linkages). The distances provide for a measure of similarity between objects and may be measured in a variety of ways, such as Euclidean and Manhattan metric distance. The criteria used to then link (group) the variables may also be undertaken in a variety of manners, as a result significant variation in results may be seen. Linkages are based upon how the association between groups is measured. For example, simple linkage or nearest neighbor distance, measures the distance to the nearest object in a group while furthest neighbor linkage or complete linkage, measures the distance between furthest objects. These linkages are both based upon single data values within groups, whereas average between group linkages is based upon the distance from all objects in a group. Centroid linkage has a new value, representing the group centroid, which is compared to the ungrouped point to weigh inclusion. Ward's method is variance based with the groups variance assessed to enable clustering. The group which sees the smallest increase in variance with the iterative inclusion of a case will receive the case. Ward's is a popular default

linkage which produces compact groups of well distributed size. Standardization of variables is undertaken to enable the comparison of variables to minimize the bias in weighting which may result from differing measurement scales and ranges. Z score format accounts for differences between mean values and reduces the standard deviation when variables have multivariate normality

9.6.2 Choosing number of groups:

The ideal number of groups to establish may be assessed graphically or numerically. Graphically the number of groups may be assessed with an icicle plot or dendrogram. The dendrogram bisected at a point which will divide the cases into a cluster based upon groupings up to the point where the bisection occurred. Numerically the number of cases may be assessed on the agglomeration schedule, by counting up from the bottom to where a significant break in slope (numbers) occurs. The optimal number of groups may be assessed based upon knowledge of the data set. Discriminant analysis may also be employed to assess optimality and efficiency of computed groups, by imputing the cluster analysis derived classes for analysis with the original data. Like the other techniques, cluster analysis presents the problem of how many factors, or dimensions or clusters to keep. One rule followed here is to choose a place where the cluster structure remains stable for a long distance. Some other possibilities are to look for cluster groupings that agree with existing or expected structures, or to replicate the analysis on subsets of the data to see if the structures emerge consistently.

The purpose of cluster analysis is to reduce a large data set to meaningful subgroups of individuals or objects. The division is accomplished on the basis of similarity of the objects across a set of specified characteristics. Outliers are a problem with this technique, often caused by too many irrelevant variables. The sample should be representative of the population, and it is desirable to have uncorrelated factors. There are three main clustering methods: hierarchical, which is a

treelike process appropriate for smaller data sets; nonhierarchical, which requires specification of the number of clusters a priori, and a combination of both. There are 4 main rules for developing clusters: the clusters should be different, they should be reachable, they should be measurable, and the clusters should be profitable (big enough to matter). This is a great tool for market segmentation. Cluster analysis is an exploratory tool designed to reveal natural groupings within a large group of observations. Cluster analysis segments the survey sample, i.e. respondents or companies, into a small number of groups. Respondents whose answers are very similar should fall into the same clusters while respondents with very different answers should be in a different cluster. Ideally, the cases in each group should have a very similar profile towards specific characteristics (e.g. attitudinal or behavioural questions), while the profiles of respondents belonging to different clusters should very dissimilar. Its main advantage is that it can suggest, based on complex input, groupings that would not otherwise be apparent ie the needs of specific groupings or segments in the market.Cluster analysis is widely used in market research to describe and quantify customer segments. This enables marketers to target customers tailored to their needs instead of having one general marketing approach - see market segmentation

9.7 CORRESPONDENCE ANALYSIS

Correspondence analysis is a technique which allows rows and columns of a data matrix, e.g. average satisfaction scores for several products, to be displayed as points in a two-dimensional space or map. It reduces a complicated set of data to a graphical display which is immediately and easily interpretable. Brand maps are based on correspondence analysis. Brand maps are often used to illustrate customers' images of the market by placing products and attributes together on a map. This allows close interpretation of company perceptions with a variety of product and service attributes simultaneously. Brands are most strongly associated with the attributes that are closest to them on the map. If products are

placed close to each other, it means they have a similar image or profile in the market. The relative association of brands with an attribute can be determined by drawing a perpendicular line from the attribute vector line (=line from the origin to the attribute point) to each of the brands. The distance between the brand and the attribute is the distance between the attribute location and where the perpendicular line crosses the attribute vector line. The centre of the map (the cross on the map), represents the overall mean of each attribute, and is the centre around which the brands are dispersed. The more a brand tends to lie in a similar direction away from the centre as an attribute, the more a brand is associated with that attribute. This also means that brands and attributes near the centre of the maps are not differentiating. The length of an attribute vector represents the extent to which the brands differ on that attribute. Angles between the vectors represent correlations between attributes. The smaller the angles, the more correlated the attributes are.

9.8 CONJOINT ANALYSIS

This technique provides for dimensional reduction of object ratings on a set of attributes, resulting in a perceptual map of the ratings. However, unlike MDS, both independent variables and dependent variables are examined at the same time. This technique is more similar in nature to factor analysis. It is a compositional technique, and is useful when there are many attributes and many companies. It is most often used in assessing the effectiveness of advertising campaigns. It is also used when the attributes are too similar for factor analysis to be meaningful. The main structural approach is the development of a contingency (crosstab) table. This means that the form of the variables should be nonmetric. The model can be assessed by examining the Chi-square value for the model. Correspondence analysis is difficult to interpret, as the dimensions are a combination of independent and dependent variables. Conjoint analysis is often referred to as "trade-off analysis," in that it allows for the evaluation of objects and the various levels of the attributes to be examined. It is both

Check your Progress

5) What is conjoint analysis?

a compositional technique and a dependence technique, in that a level of preference for a combination of attributes and levels is developed. A part-worth, or utility, is calculated for each level of each attribute, and combinations of attributes at specific levels are summed to develop the overall preference for the attribute at each level. Models can be built which identify the ideal levels and combinations of attributes for products and services. Market research is frequently concerned about finding out which aspects of a product or service is most important to companies. The ideal product or service, of course, would have all the best characteristics, but realistically, trade-offs have to be made. The product with the most expensive features, for example, cannot have the lowest price. Conjoint analysis is a technique for measuring respondent preferences about the attributes of a product or service. It is the ideal tool for new/improved product development. The conjoint analysis task asks the respondents to make choices in the same fashion as consumers normally do, by trading off features one against the other, either by ranking or choosing one of several product combinations. e.g. a task could be: do you prefer a "flight that is cramped, costs Rs.25000 and has one stop" or a "flight that is spacious, costs Rs.50000 and is direct"? Using conjoint analysis, you can determine both the relative importance of each attribute (e.g. spaciousness, price, number of stops) as well as which levels of each attribute are most preferred (e.g. how much is a price of Rs.25000 more preferred than a price of Rs.50000). Example: Importance Of Printer Features, Plus Simulator Conjoint analysis is typically used to guide new product developers by indicating which product aspects are most important to different companies. It is also useful to gauge market reaction when a product (attribute) will change. e.g. what will happen to the market share of brand A if its price increases with 10%?

Conjoint analysis deals with the measurement of the combined effect of two or more attributes that are important from the view of the consumer. The use of the conjoint analysis will be more appropriate in a situation

where a company would like to know the most desirable attribute for a new product or service. For example, a hotel would like to know whether choice of menu or prompt service would attract a customer to visit them frequently. For this, it will seek data from the consumer in the firm of response to identify product attributes. The various options available for this are direct interview with the customer or focus group interviews. All the attributes are weighed and compared. The main steps involved in the application of conjoint analysis are

1. Determination of salient attributes: the attributes have to be selected based on the marketers experience or through interviews. Only valuable attributes need to be considered.
2. Assigning levels to the selected attributes: this can vary from most preferred to the least preferred.
3. Fractional factorial design of experiments: during the comparison of the profile of different products, it is essential to have a minimum number of designs which provide us all the information required. This will ensure easy management of the design.
4. Physical design of the stimuli: a prototype or a picture of the concept may be given to the consumer or customer to get a realistic picture.
5. Data collection: The customers are asked to rank all the alternatives using a rating scale. This will ensure ease of data collection and analysis.
6. Determination of part-worth utilities: Regression methods, mathematical programming methods, econometric methods may be used for the part-worth utility values. The applications of conjoint analysis are

- It can be used for optimum product design based on the attributes considered. Simulations can be created to represent competitors' action or a fresh scenario.
- Consumers can be segmented based on their sensitivity to product attributes.
- It may help a manager to conduct SWOT analysis of the brand as the part-worth utility speaks about the relative brand strength. There are

certain limitations of conjoint analysis. It may not be perfect and convincing and may fail to capture utility functions and decision roles.

9.9 CHAID ANALYSIS

CHAID (Chi Squared Automatic Interaction Detection) is used to build a predictive model, based on a classification system. The analysis subdivides the sample into a series of subgroups that 1) share similar characteristics towards a specific response variable and that 2) maximises our ability to predict the values of the response variable. The first predictor category (on which the sample will be split) is the predictor that is associated the most with the response variable. i.e. it gives the most differentiating groups of respondents. Each group is then further split until the analysis does not find any significantly discriminating predictor any more. The predictors can be scaled (e.g. 1 to 10 scale rating) as well as categorical questions (e.g. company demographics). The output is a tree of which the branches are the predictor variables that split the sample in discriminating groups. CHAID is very often used to understand the characteristics of the most and least satisfied or interested customers/employees. It allows the client to target its (potential) clients more efficiently and successfully. CHAID analysis is typically used in the direct marketing industry to identify the type of people who have reacted to a specific campaign.

9.10 DISCRIMINANT ANALYSIS

The purpose of discriminant analysis is to correctly classify observations or people into homogeneous groups. The independent variables must be metric and must have a high degree of normality. Discriminant analysis builds a linear discriminant function, which can then be used to classify the observations. The overall fit is assessed by looking at the degree to which the group means differ (Wilkes Lambda or D2) and how well the model classifies. To determine which variables have the most impact on the discriminant function, it is possible to look at partial F values. The higher the partial F, the more impact that variable has on the

discriminant function. This tool helps categorize people, like buyers and nonbuyers. Discriminant and logistic regression analysis are statistical techniques that point out the differences between two or more groups based on several characteristics (most often rating scales when discriminant analysis, while logistic regression can handle any type of variable). It explains why respondents belong to a certain group, plus it classifies new respondents based on their ratings. i.e. it explains and predicts classification. e.g. why are people very satisfied (i.e. gave a score 9 or 10 on a 1-10 point scale) with a product versus the rest of the market? Is often used to determine which customers are likely to buy a company's product to decide whether a bank should offer a loan to a new company or to identify patients which may be at high risk for medical problems.

Discriminant analysis is used to classify the sample into two or more categories. Example: Consumers may be classified as heavy and light users; Sales people can be classified as successful and unsuccessful and so on. Discriminant function analysis is used to determine which variables discriminate between two or more naturally occurring groups. For example, a researcher may want to investigate which variables discriminate between engineers who decide

(1) To seek employment in private companies,

(2) To take up government services, or

(3) To seek opportunities abroad.

For that purpose the researcher could collect data on numerous variables after the graduation of the engineers. *Discriminant Analysis* could then be used to determine which variable(s) are the best predictors of the engineers' choice of employment. The objectives of two group discriminant analysis are to find a linear composite of the predictor variable to help the analyst to separate the groups, establishing procedures for assigning new individuals, testing for significant differences between the mean predictor variables and determining the variable which accounts

for the most intergroup differences. This is commonly carried out with the help of a computer program.

9.11 MULTIDIMENSIONAL SCALING

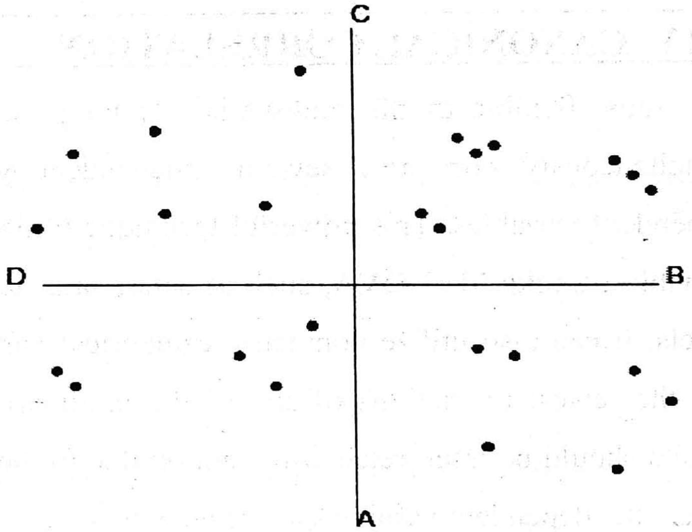
Multidimensional scaling (MDS) can be considered to be an alternative to factor analysis. In general, the goal of the analysis is to detect meaningful underlying dimensions that allow the researcher to explain observed similarities or dissimilarities between the investigated objects. In factor analysis, the similarities between objects (e.g. variables) are expressed in the correlation matrix. With MDS one may analyze any kind of similarity or dissimilarity matrix, in addition to correlation matrices. This outcome is visualised in a 2 dimensional map, which gives the researcher an immediate feel of how differentiating the questions were. Questions which are clustered together did get very similar scores by all respondents. This can be very useful when optimising a questionnaire or to differentiate consumers based on the most distinct questions. Even though there are similarities in the type of research questions to which MDA and factor analysis can be applied, they are fundamentally different methods. Factor analysis requires that the underlying data are distributed as multivariate normal, and that the relationships are linear. MDS imposes no such restrictions. Just as long as the rank-ordering similarities in the matrix are meaningful, MDS can be used.

In terms of resultant differences, factor analysis tends to extract more factors (dimensions) than MDS; as a result, MDS often yields more readily, interpretable solutions. Most importantly, however, MDS can be applied to any kind of similarities, while factor analysis requires us to first compute a correlation matrix. MDS can be based on subjects' direct assessment of similarities between stimuli, while factor analysis requires subjects to rate those stimuli on some list of attributes (for which the factor analysis is performed). In summary, MDS methods are applicable to a wide variety of research designs.

NOTES

Multi-dimensional scaling (MDS) or perceptual map or positioning map is used for measuring human perception and preferences. It is spatial representation of relationships. It helps in the identification of attributes and the positioning of different products or brands on the basis of these attributes. MDS is of two types, metric MDS and non metric MDS. Given below is an example of a perceptual map. AC represents the x-axis and BD represents the y-axis. The value of the variable may be low to high from one end of the axis to the other. Two approaches can be used for analyzing multi dimensional data. It can be done by measuring the attributes or distance between objects. For MDS, a set of number called proximities and a computer based algorithms must be available. The applications of MDS in marketing are in market segmentation, vendor evaluation, attitude scaling, advertisement evaluation, product repositioning, new product development and test marketing. Thus MDS measures the psychological distance or the dis-similarities to evaluate the external environment. The limitations of MDS are

- Concepts of similarity and preferences may differ
- The selection of attributes are subjective
- It is at time difficult to interpret the results
- Different computer programmes may produce different results



9.12 STRUCTURAL EQUATION MODELING

Unlike the other multivariate techniques discussed, structural equation modeling (SEM) examines multiple relationships between sets of variables simultaneously. This represents a family of techniques, including LISREL, latent variable analysis, and confirmatory factor analysis. SEM can incorporate latent variables, which either are not or cannot be measured directly into the analysis. For example, intelligence levels can only be inferred, with direct measurement of variables like test scores, level of education, grade point average, and other related measures. These tools are often used to evaluate many scaled attributes or build summated scales. Structural Equation Modeling (SEM) is a very general, very powerful multivariate analysis technique that includes a number of other traditional analysis methods as special cases. It effectively includes a whole range of standard multivariate analysis methods, such as regression, factor analysis and analysis of variance. A structural equation model can exist with several regression and factor analysis models, which are estimated simultaneously. It is a statistical methodology that takes a hypothesis-testing (i.e. confirmatory) approach to the multivariate analysis. SEM tests a theory using survey data, while traditional modeling uses the data to build a model (i.e. exploratory approach). Commonly used for validating models, e.g. a CRM model using survey.

9.13 CANONICAL CORRELATION

The most flexible of the multivariate techniques, canonical correlation simultaneously correlates several independent variables and several dependent variables. This powerful technique utilizes metric independent variables, unlike MANOVA, such as sales, satisfaction levels, and usage levels. It can also utilize nonmetric categorical variables. This technique has the fewest restrictions of any of the multivariate techniques, so the results should be interpreted with caution due to the relaxed assumptions. Often, the dependent variables are related, and the independent variables

are related, so finding a relationship is difficult without a technique like canonical correlation.

9.14 ANALYSIS OF VARIANCE

This technique examines the relationship between several categorical independent variables and two or more metric dependent variables. Whereas analysis of variance (ANOVA) assesses the differences between groups (by using T tests for 2 means and F tests between 3 or more means), MANOVA examines the dependence relationship between a set of dependent measures across a set of groups. Typically this analysis is used in experimental design, and usually a hypothesized relationship between dependent measures is used. This technique is slightly different in that the independent variables are categorical and the dependent variable is metric. Sample size is an issue, with 15-20 observations needed per cell. However, too many observations per cell (over 30) and the technique loses its practical significance. Cell sizes should be roughly equal, with the largest cell having less than 1.5 times the observations of the smallest cell. That is because, in this technique, normality of the dependent variables is important. The model fit is determined by examining mean vector equivalents across groups. If there is a significant difference in the means, the null hypothesis can be rejected and treatment differences can be determined. This can be studied under ANOVA, ANCOVA, MANOVA and MANCOVA.

9.14.1 ANOVA: An ANOVA (Analysis of Variance), sometimes called an F test, is closely related to the t test. The t test measures the difference between the means of two groups whereas an ANOVA tests the difference between the means of two or more groups. ANOVA can be a one-way ANOVA or a factorial ANOVA. One-way or single factor ANOVA, tests differences between groups that are only classified on one independent variable. A factorial ANOVA can show whether there are significant main effects of the independent variables and whether there are significant interaction effects between independent variables in a set of data.

Interaction effects occur when the impact of one independent variable depends on the level of the second independent variable. The advantage of using ANOVA rather than multiple t-tests is that it reduces the probability of a type-I error. Making multiple comparisons increases the likelihood of finding something by chance—making a type-I error. An F indicates that there is a significant difference between groups, not which groups are significantly different from each other. This is one potential drawback to an ANOVA, which is loss of specificity. A factorial ANOVA can examine data that are classified on multiple independent variables. More than two independent variables can be compared in an ANOVA (e.g., three-way, four-way).

9.14.2 ANCOVA: In ANCOVA, we can analyze both qualitative (class) and quantitative (continuous) independent variables. The mixed procedure allows the user to model both class and continuous variables. In ANOVA-type models, hypotheses about class or interactions among class variables are tests of means or differences among means. In regression-type models, in which all "factors" are continuous variables (rather than categories), hypothesis tests are tests about regression coefficients. As one might expect, the assumptions of ANCOVA combines both the assumptions of regression and ANOVA. In addition, the tests of adjusted means are based on the assumption that the class variable by covariate interaction is negligible, that is the regression lines are parallel.

9.14.3 MANOVA: Multivariate analysis of variance (MANOVA) is simply an ANOVA with several dependent variables. For example, we may conduct a study where we try two different textbooks, and we are interested in the students' improvements in Physics and Chemistry. In that case, improvements in physics and chemistry are the two dependent variables, and our hypothesis is that both together are affected by the difference in textbooks. A multivariate analysis of variance (MANOVA) could be used to test this hypothesis. Instead of a univariate F value, we would obtain a multivariate F value.

"covariance" here is included because the two measures are probably correlated and we must take this correlation into account when performing the significance test. MANOVA is useful in experimental situations where at least some of the independent variables are manipulated. It has several advantages over ANOVA.

1. By measuring several dependent variables in a single experiment, there is a better chance of discovering which factor is truly important.

2. It can protect against Type I errors that might occur if multiple ANOVA's were conducted independently. Additionally, it can reveal differences not discovered by ANOVA tests.

However, there are several cautions as well. It is a substantially more complicated design than ANOVA, and therefore there can be some ambiguity as to which independent variable affects each dependent variable. Moreover, one degree of freedom is lost for each dependent variable that is added. Finally, the dependent variables should be largely uncorrelated. If the dependent variables are highly correlated, there is little advantage in including more than one in the test given the resultant loss in degrees of freedom. Some of the assumptions made here are:

• Normal Distribution: The dependent variable should be normally distributed within groups. Overall, the F test is robust to non-normality if it is caused by skewness rather than outliers. Tests for outliers should be run before performing a MANOVA, and outliers should be transformed or removed.

• Homogeneity of Variances: Homogeneity of variances assumes that the dependent variables exhibit equal levels of variance across the range of predictor variables.

• Homogeneity of Variances and Covariance's: In multivariate designs, with multiple dependent measures, the homogeneity of variances assumption described earlier also applies. However, since there are multiple dependent variables, it is also required that their covariances are

homogeneous across the cells of the design. There are various specific tests of this assumption.

Two Special Cases arise in MANOVA:

• Unequal sample sizes: As in ANOVA, when cells in a factorial MANOVA have different sample sizes, the sum of squares for effect plus error does not equal the total sum of squares. This causes tests of main effects and interactions to be correlated.

• Within-subjects design: Problems arise if the researcher measures several different dependent variables on different occasions.

9.14.4 MANCOVA: MANCOVA is an extension of ANCOVA. It is simply a MANOVA where the artificial direct variables are initially adjusted for differences in one or more covariates. This can reduce error "noise" when error associated with the covariate is removed.

9.15 LOGISTIC REGRESSION ANALYSIS

Sometimes referred to as "choice models," this technique is a variation of multiple regression that allows for the prediction of an event. It is allowable to utilize nonmetric (typically binary) dependent variables, as the objective is to arrive at a probabilistic assessment of a binary choice. The independent variables can be either discrete or continuous. A contingency table is produced, which shows the classification of observations as to whether the observed and predicted events match. The sum of events that were predicted to occur which actually did occur and the events that were predicted not to occur which actually did not occur, divided by the total number of events, is a measure of the effectiveness of the model. This tool helps predict the choices consumers might make when presented with alternatives.

9.16 SUMMARY

This unit deals with the different methods used in multivariate analysis and briefly describes their application. The methods are chosen based on the data and requirements of analysis. Each of the multivariate techniques described above has a specific type of research question for

which it is best suited. Each technique also has certain strengths and weaknesses that should be clearly understood by the analyst before attempting to interpret the results of the technique. Current statistical packages (SAS, SPSS, S-Plus, and others) make it increasingly easy to run a procedure, but the results can be disastrously misinterpreted without adequate care.

9.17 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Multivariate analysis is the analysis of t he simultaneous relationships among three or more phenomena.

2. R^2 denotes the power of regression.

3. Regression is the estimation of an unknown variable normally referred to as dependent variable with the help of a known variable referred to as independent variable.

4. Factor analysis is a technique with the purpose of data reduction and summarization.

5. Conjoint analysis is a technique that provides for dimensional reduction of object ratings on a set of attributes, resulting in a perceptual map of the ratings.

9.18 QUESTIONS

1. What is multivariate analysis? Explain.
2. Briefly explain the methods of multivariate analysis.
3. Explain variance of analysis. Write a note on the different types.
4. What is multiple regression analysis?
5. What is discriminant analysis? How is it useful in Marketing?
6. What are the steps involved in the application of conjoint analysis?
7. What are the limitations and uses of factor analysis?
8. What is cluster analysis? What are the steps involved in cluster analysis?
9. What is Multi-dimensional scaling? Where is it used?
10. How do you choose a method for analysis?

NOTES

9.19 FURTHER READING

1 Boyd Harper W., Ralph Westfall and Stanley F Stasch, *Marketing Research Text and Cases*, Irwin, Inc., 2007.

2 G.C.Beri, *Marketing Research*, TMH, 2008.

3 Nargundkar R, *Marketing Research Text and Cases,* TMH, 2002

4 Paul. E. Green, Donald. S. Tull, Gerald Albaum, *Research for Marketing Decisions*, Prentice Hall of India Pvt. Ltd. New Delhi.

5 Dr.D.D.Sharma, *Marketing research, Principles, applications, and cases,* Sultan Chand and sons, 2007.

NOTES

UNIT X

APPLICATIONS OF MARKETING RESEARCH

STRUCTURE

10.1 Introduction
10.2 Product Research
10.3 Price Research
10.4 Distribution Research
10.5 Promotion Research
10.6 Advertising research
10.7 Branding research
10.8 Package research
10.9 Segmentation research
10.10 Summary
10.11 Answers to 'Check Your Progress'
10.12 Exercises and Questions
10.13 Further Readings

UNIT OBJECTIVES

By reading this unit you would be able to understand what is

Product Research

Price Research

Distribution Research

Promotion Research

Package research

Segmentation research

10.1 INTRODUCTION

As we have covered various theoretical concepts in marketing research now we move to its real applications.

Check your Progress

1) What are the recent applications of marketing research?

10.2 PRODUCT RESEARCH

A managerial decision to use a pretest market analysis is justified if sufficiently accurate predictions can be achieved, the timing of the analysis is before large investment commitments are necessary, useful diagnostics for improvement are generated, and the cost of the analysis is reasonable. In these situations failures can be reduced, time-to-market can be shortened, and products improved to increase customer satisfaction

10.2.1 New Product Research

New product development is critical to the life of most organizations as there will be uncertainties associated with them. Thus, the purpose of marketing research for them would reduce the uncertainties associated with the new products. The various research areas in new product research are:

1. New Product Variation - Manufacturers and service providers need to constantly look for new concepts a part of their market strategy. This helps not only to stay ahead of competition but also for maintaining and generating new markets.

2. Screening - A new product variation is developed through Research & Development. Here, the product already exists and we look for different attributes to add on new ideas to make the product more viable to the market in the form of product extensions. It could involve changing a component of the product and offering it as a new product.

3. Concept Testing - this is done by

a. Focus Groups

b. Mail Intercepts (Stop people at a mall and use them as samples)

NOTES

4. Business and Market Analysis- this is done by
a. Company audit - Introspection within the company to see whether it is meeting its sale target etc.
b. Market Testing- Testing competitive ideas, market awareness, market recognition and how the competition is doing.
5. Actual Product Development- This test is to see whether the end product does what the screen test says it will do check whether attributes in the product match the claims that have been made about it.
6. Market Test- Test market the product using one set of demographics and geographic area.
7. Commercialization of Product- The company and the ad agency must develop and evaluate all avenues of advertising and promotion.
8. Innovation, Diffusion, and Adoption- Innovation: - Introduce product variation Diffusion: - Market penetration (e.g. Surf Excel talking about water saving) Adoption: - Adjust to market changes (e.g. Jumbo Wadapav, Suzuki Swift)

10.2.2 Objectives of new product research

1) Develop overall strategy based on market trends.
2) Develop a flavor of new product ideas from a variety of sources.
3) Develop preliminary procedures for screening new ideas.
4) Develop procedure for final screening
5) Develop product specifications with regard to optimum product attributes
6) Test the product.
7) Test market the product.
8) Commercialize and supervise the product through its life cycle and its termination or phase out.

Check your Progress

2) What are the emerging applications of marketing research?

NOTES

10.2.3 Product Testing Methods

1) Paired comparison test:

Consumers are not told about the brand and are given new and an old product and are asked to choose. The sampl know that they are testing and are aware therefore, about probable difference (very much like psychological testing)

2) Staggered comparison test:

Respondents test 2 brands with a time lag with th identities masked. One half of the respondents receive brand A the other half brand B. Further, the respondents are given th same products in reverse and are asked to note any differenc in the Brands.

3) Disguised comparison Test:

This duplicates the actual market where for differen brands, the same packaging is used for different brands and the respondents are told that they will be asked about their preferences later. In this case, the respondents are studied in order to recognize if they have noticed any difference at all Further, they are not aware of the test and believe the product to be a free sample.

	PAIRED	STAGGERED	DISGUISED
Slightly Aware of difference	62	60	2
Quite Aware of difference	17	18	5
Not Aware of differences	21	22	93
	100	100	100

This study of any given product of a brand examines how disguised packaging shows that there is no or minimum recognition of difference. While respondents participating in paired comparison with those participating using the same

product in a disguised comparison test. What is interesting to note that as samples are aware of being participants in a test, even slight changes are noticed by them especially 'paired' and 'staggered' participants However, those that are not aware of being part of a test find it almost impossible to note subtle changes.

10.2.4 Predicting Trial Purchase

To predict trial levels of new, frequently purchased consumer products, ESP (Estimating Sales Potential) model has been developed. Trial levels were predicted on the basis of three variables:

- Product class penetration (PCP), the percentage of households purchasing at least one item in the product classwithin one year.
- Promotional expenditures-total consumer-directed promotional expenditures on the product.
- Distribution of the product-percentage of stores stocking the product (weighted by the store's total sales volume). Once the model is estimated, it can be applied to other new products. The researcher simply estimates the percentage of household using the product class, the total expenditures planned for the new product, and the expected distribution level. The model will then estimate the trial level that will be obtained. Trial also can be estimated directly using controlled shopping experience. A respondent is exposed to the new product promotion and allowed to shop in a simulated store or in an actual store in which the product is placed. The respondents then have an opportunity to make a "trial" or first purchase of the product.

10.2.5 Pre-testing

Functions of Pre-testing are

· To gain information and experience with the marketing program before making a total commitment to it.

· To predict the program's outcome when it is applied to the total market.

10.2.6 Types of Test Market

a. The sell-in test markets are cities in which the product is sold just as it would be in a national launch. The product has to gain distribution space.

b. The controlled-distribution scanner markets (CDSM) are cities for which distribution is pre-arranged and the purchases of a panel of customers are monitored using scanner data. Certain parameters that have to be looked into while deciding sell-in test market:

- Representativeness: Ideally, the city should be fairly representative of the country in terms of characteristics that will affect the test outcome, such as product usage, attitudes and demographics.
- Data Availability: Information about Store audit is helpful in evaluating the test. The selected cities should contain retailers who will cooperate with store audits.
- Media isolation and Costs: It is desirable to avoid media spill-over. Using media that "spill-out" into nearby cities is wasteful and increases costs. Conversely, "spill-in" media from nearby cities can contaminate a test. Media cost is another consideration.
- Product flow: It may be desirable to use cities that don't have much "product-spillage" outside the area. · Number: A single city can lead to unreliable results because of the

variations across cities of both brand sales and consumer response to marketing programs.

➢ Implementing and controlling: The test should be controlled in such a manner that it ensures the marketing program is implemented in the test area so as to reflect the national program. The test itself may tend to encourage those involved to enhance the effectiveness of the marketing program. Salespersons may be more aggressive. Retailers may be more cooperative. The competitors may react by deliberately flooding the test areas with free samples or instore promotions. Even they can retaliate or can also monitor the results themselves.

➢ Timing: Normally, a test market should be in existence for one year, so that all important seasonal/cultural factors can be observed and estimated.

➢ Measurement: The basic measure is sales based on shipments or warehouse withdrawals. Store audit data provide actual sales figures and are not sensitive to inventory fluctuations. They also provide information on: distribution, shelf-facings, and in-store promotional activity. Measures such as brand awareness, attitude, trial purchase, and repeat purchase are obtained directly from the consumer. This information helps evaluate the marketing program and can help interpret sales data. The most useful information obtained from consumers is whether they bought the product at least once, whether they were satisfied with it, and whether they repurchased it or plan to.

➢ Costs: Costs which are quantifiable, include – development and implementation of the marketing program; preparation of test products; administration of the test and collection of data associated with the test. The costs and risks that may delay the

➤ Launch of a new product are more difficult to quantify. If a new product launch is delayed, an opportunity to gain a substantial market position might be lost.

10.3 PRICING RESEARCH

Research may be used to evaluate alternative price approaches for new products before launch or for proposed changes in products already in the market.

10.3.1 Pricing Approaches

These are 3 main approaches to pricing research.

1. Gabor and Grainger Method (Price skimming strategy), where different prices for a product are presented to respondents, who then are asked if they would buy. A "buy response" curve of different prices, with the corresponding number of affirmative purchase intentions, is produced. The objective is to generate as much profit as possible in the present period. Gabor Granger research is named after the economists who invented it in

the 1960's. The customers are 1st asked if they would buy a product at a particular price, the price is then changed and once again the respondents are asked the same question. By taking sample responses of customers and analyzing them researchers can see what is the level of demand at each price level.

2. Conjoint Analysis - It is a technique that allows to workout the hidden rules that people use to differentiate between products and services by understanding precisely how people decide and what you can workout, what are the features and services in comparison with the cost that the customers are looking far.

3. Multi brand-choice Method (Share penetration strategy), where respondents are shown different sets of brands in the same product category, at different prices, and are asked which

Check your Progress

3 What is the role of market research in a product's life cycle?

they would buy. This allows the respondents to take into account competitors' brands, as they normally would outside such a test. Thus, this technique represents a form of simulation of the point of sale. The objective is to capture an increasingly larger market share by offering a lower price. Pricing research for the two different approaches differs substantially in terms of the information sought. Following questions are generally asked with regard to pricing research:

1. At what price would you consider the product to be so expensive that you would not consider buying it? (Too expensive)
2. At what price would you consider the product to be priced so low that you would feel the quality couldn't be very good? (Too cheap)
3. At what price would you consider the product starting to get expensive, so that it is not out of the question, but you would have to give some thought to buying it? (Expensive)
4. At what price would you consider the product to be a bargain-a great buy for the money? (Cheap) Research for Skimming Pricing This is based on the concept of pricing the product, at the point at which profits will be the greatest until market conditions change or supply costs dictate a price change. Under this strategy, the optimal price is the one that results in the greatest positive difference between total revenues and total costs. This implies that the researcher's major tasks are to forecast the costs and the revenues over the relevant range of alternative prices.

10.3.2 Research for Penetration Pricing

This is based on the concept that average unit of production costs continue to go down as cumulative output increases. Potential profits in the early stages of the product life cycle are

sacrificed in the expectation that higher volumes in later periods will generate sufficiently greater profits to result in overall profit for the product over its life. Following pricing pattern is adopted to increase market share:

a. Offer a lower price (even below cost) when entering the market.

b. Hold that price constant until unit costs produce a desired percentage markup.

c. Reduce price as costs fail to maintain markup at the same desired percentage.

10.4 DISTRIBUTION RESEARCH

Traditionally, the distribution decisions in marketing strategy involve:

- The number and location of salespersons,
- Retail outlets,
- Warehouses, and
- The size of discount to be offered.

The discount to be offered to the members in the channel of distribution usually is determined by what existing or similar products are offered, and also whether the firm wants to follow a "push" or a "pull" strategy.

10.4.1 Warehouse and Retail Location Research

Location decisions include: "What costs and delivery time would result, if we choose one location over another?" The approximate location (optimal location), that will minimize the distance to customers, weighted by the quantities purchased, will have to be determined. Chain shops with multiple outlets and franchise operations must decide on the physical location of their outlets. Data about surrounding residential neighbourhood, income levels, and competitive stores would help in choosing optimal location. Number and

location of Sales Representatives How many sales representatives should be there in a given territory?

Approaches

a. *Sales effort approach*- when the product line is first introduced and there is no operating history to provide sales data. This is done by: Estimating the number of sales calls required to sell to, and to service, prospective customers in an area for a year. This will the sum of the number of visits required per year to each prospect (customer) in the territory. Estimating the average number of sales calls per representative that can be made in that territory. Divide the estimate in step (i) by the estimate in step (ii) to obtain the number of sales representatives required.

b. *Statistical analysis approach*- is used after the sales program is under way. Once a sales history is available from each territory. an analysis can be made to determine if the appropriate number of sales representatives is being used ineach territory. An analysis of actual sales versus market potential for each sales representative can be made. Also, following inferences can be made:

i. Average market potential is less as per each sales representative

ii. Territory, which have too many sales representatives

iii. Market potential is more but have too few sales representatives

c *Field Experiment approach*- is also applicable only after the sales program has begun. Experiments are done with the calls made, to determine the number and location of sales representatives. This is done in two ways:

i. Making more frequent calls on some prospects and less frequent calls on others, in order to see the effect on overall sales, keeping the number of sales representatives unchanged.

ii. Increasing the number of representatives in some territories and decreasing them in others to determine the sales effect.

10.6 PROMOTION RESEARCH

Here the focus is on the decisions that are commonly made when designing a promotion strategy. The decision for the promotion part of a market strategy can be divided into:

· Advertising decisions, which have long-term effects.

· Sales Promotion decisions, which affect the company in the short term. Companies spend more time and resources on advertising research than on sales promotion research because of the greater risk and uncertainty in advertising research.

10.7 ADVERTISING RESEARCH

Advertising research is a specialized form of marketing research conducted to improve the efficiency of advertising. According to green and tull "It may focus on a specific ad or campaign, or may be directed at a more general understanding of how advertising works or how consumers use the information in advertising. It can entail a variety of research approaches, including psychological, sociological, economic, and other perspectives. Advertising research is a branch of marketing research, and is both a sort of insurance to avoid wasting money on in effective advertising and a means of monitoring the effectiveness of a campaign while it is running and after. Advertising research helps making decisions in the:

· Awareness stage

· Recognition stage

· Preference stage, and

· Purchasing stage

Most often, advertising research decisions are about advertising copy. Marketing research helps to determine how effective the advertisement will be. Research on media decisions is separate from advertising research.

Applications of MR

NOTES

10.7.1 Copy Testing – Measures & Methods

To test and copy, the researches must know both what the copywriter is trying to accomplish and what assumptions are being made as to how the various copy components will contribute to this end. Typical measures used to judge advertising effectiveness Measures dealing with recognition, recall, comprehension, believability, persuasion and attitude change. All involved assumptions on how advertising works.

a. Pre-Test - It is used to use to determine what weaknesses exist in the copy and concept before too much money time, marketing, schedules and plans have been put in place on the advertising.

b. Post-Test - It attempts to measure the combined effect of the advertising, the media used, the scheduling, the products distribution and competitive advertising.

c. Creative Strategy Research - A products creative strategy is concerned mainly with determining what message can best elicit the desired response from members of the target audience. Consumer Jury Personal interviews are used or a group is assembled and the members asked to vote on the alternatives. It provides a rating given to an advertisement by a group of consumers who represent potential buyers of the product.

d Portfolio Test - These tests are named after the manner in which the advertisements to be tested are packaged. The ads are placed in a port folio. The respondents are asked to

recall with the folio closed, the ads that they can remember. Such recall may be on a completely unaided basis, or the interviewer may aid recall by asking about specific ads or ads for specific products. For each recalled ad the respondent is asked to playback as much of the ad as possible. The information is recorded verbatim.

e. Psychological Tests - They differ in the methods used Advertising effectiveness depends on the results achieved in mind of the individual reached by the advertisement. Ideally, one can list reactions such as: - self pity, security, fear, nostalgia, anxiety etc. Psychological tests employ clinical research techniques such as words association, sentence completion, in depth interviews, and story testing. Only skilled interviews can be used.

f. Inquiries [Post Test]- It measures the advertisements worth.

1) Place the ads in different places in a copy where all other factors remain equal then check how man inquiries are made.

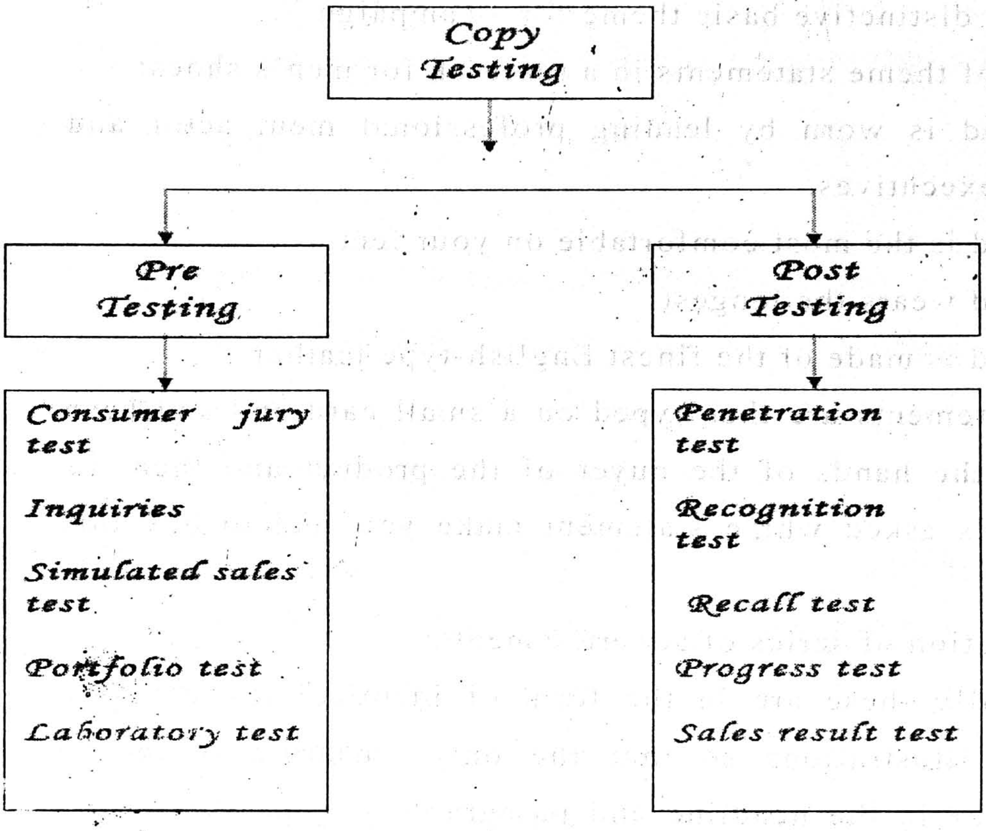
2) Place the same ad in different magazines and do a similar check.

3) Split – run

g. Sales Tests [Post Test] - These tests are done by either using POP's or direct mail.

1. POP Test - Consumers are exposed to alternative pieces of copy or product sales are measured into stores keeping the same copy.

2. Direct Mail Test - This is done with the use of coupons. One group is shown a TV ad and the other is not. Then both groups are given coupons, to buy the product that has been advertised. The researcher then measures the influence of TV ads on both groups.



10.7.2 Pre-Testing

Pre-testing attempts to measure the communication effect of an advertisement before it is released in the media. Various tests conducted under this category are as follows:

a. Consumer Jury Tests - The consumer jury test is based on consumer's preferences under controlled conditions in which they are presented with alternative choices. The consumer rates themes or advertisements according to his opinion by direct comparison. This type of rating, fundamental to much experimentation in psychology, has been effectively adapted to advertising copy research. In using the consumer jury method to evaluate advertising themes, two methods are most commonly used

1. Theme preference technique:

It involves the preparation of series of statements regarding the product. Each statement

presents a distinctive basic theme for a campaign.

Example of theme statements in a research for men's shoes:

This brand is worn by leading professional men, actor and business executives.

This brand is the most comfortable on your feet.

This brand wears the longest.

This brand is made of the finest English-type leather.

These statements are then typed on a small card and and been given in the hands of the buyer of the product and then the question is asked which statement make you feel to buy this product.

2. Preparation of series of advertisements:

Usually these are in the form of identical layouts with identical illustrations so that the only variation is in the headline or in the headline and paragraph of copy. These are then shown to consumers, who indicate their preferences. The consumer jury test is also used to evaluate various elements of advertisements. Consumers are shown two or more sample advertisements and asked which advertisement interest them the most. In conducting consumer jury test there are three aspects to be handled with utmost care:

The selection of individuals to be interviewed

The preparation of material to be used

Interpretation of the results

The consumer jury test has one outstanding advantage – it can be employed before advertisements have been run and a large sum of money spent. Since it is the most flexible of all forms, a wide variety of ideas can be tested economically. One of its greatest values is that it can warn an advertiser against some basic mistake which might be made through failure to

understand the reaction of consumers to a proposed advertising efforts.

b. Inquiries - The inquiry test refers to the number of enquiries an advertisement is able to attract after it is run on experimental basis. In this method the ads to be tested are selected. The ads carry some offer to consumers and are published one by one in different issues in some magazines or newspapers. The numbers of enquiries coming in response to those ads are measured. The difference in number of enquiries from each issue is noted and analysed. The ad with the highest number of enquiries shows the maximum attractive strength. The returned enquiries show the quality of ad. The advantage of this method is that it is very quick and systematic. Its disadvantage is that all the people making enquiries may not be the potential customers of the product advertised. Moreover the method is very expensive and needs a lot of precautions in designing the ads.

c. Simulated Sales Test - Sales test apply the experimental procedure to copy testing by making sales results the measure of the effectiveness of advertising copy. The test exposes the consumers to alternative copies of the ad through the point of purchase displays or direct mail. Indirect mail the different copies are sent to different groups of customers and effect on sales due to these groups of customers is noted. A group making maximum purchases after receipt of the ad, is supposed to carry the best copy of the advertisement. This method is simple and economical, but also it is artificial and misleading in practice, because the increase in sales at the point of display does not ensure similar increase in the sales when the ad is actually released in the media.

d. Portfolio Method - Portfolio method is a method in which a number of dummy ads are put in a folio along with the ad to be tested. The folio is given to the consumers and they are asked to see each ad in the folio. After closing the folio the consumer is asked to play back what he has seen about each ad. The ad with the maximum contents is regarded as the best. If the dummy ad is adjudged than the actual ad is improved on the same basis. This method is more theoretical than practical and this is one of its major drawbacks. But on the other side this method is simple, quick and very economical.

e. Laboratory Tests - Laboratory tests use equipment to measure physiological reactions – heartbeat, blood pressure, skin response – to an ad or consumers may be asked to turn a knob to indicate their moment to moment liking interest while viewing sequenced material. These tests measure attention getting power but reveal nothing about impact on beliefs, attitudes or intentions.

10.7.3 Post-Testing

The post testing is a technique, which not only measures effectiveness of an ad but also attempts to find why a given ad is effective or ineffective. The techniques of post-testing try to measure the memory of respondents about the ad in question. The various techniques of this testing are as follows:

a. Penetration Test - This test try to find out whether readers have read the ad and up to what extent the knowledge about the given brand has penetrated in their minds. Two types of penetration test are as follows:

b. Recognition test - In this test readers are shown issues of magazines, they claim to have read and results are tabulated. These data's are now fixed in a formula to calculate the readers per rupee.

*Readers per rupee = percent noted * magazines primary readers / Space cost*

These test are very simple and economical and simple and also create interest in consumers regarding the ads.

c. Recall Tests- This test deals with asking consumers what they can recall about he ad or what they remember about the ad. In the recall test only the qualified readers are asked so that they can describe at least one feature of the issue in which the ad in question has appeared. This test is measure that to what extent the ad has penetrated in the mind of a consumer. This test have a limited value for measuring advertising effectiveness because only qualified readers are the part of this test not all the readers, due to this reason this test has limited measure.

d. Progress Test - The progress tests try to measure the attitudes of the consumer towards products brand in question and sales result. Through progress test the advertising effectiveness is measured in terms of sales. These tests are simple and easy to execute with the minimum expenditure involved. Advertising effectiveness is measured immediately at one particular spot so results are not influenced by other factors. Awareness and preferences are there in progress tests but they may not lead to actual sales. As the consumer is asked immediately after showing the advertisement, he may be biased as he does not get time to think as in usual practice

e. Sales result test -This measures the advertising effectiveness in terms of increase in sales. Under this method one has to choose two groups matched markets with almost the same ad expenditure and sales. The least square analysis can

also be used to find a relationship between past ad expenditure and sales of previous years and sales from the current year can be predicted. This test help in estimating fairly accurately the shape of the sales response function in different types of territories. This is the best way to determine the impact of advertising expenditure on sales. But, on the other side it is not possible to find similar markets with regard to advertising expenditure and sales. Every market is different from other markets in some aspect or the other. If the competitors also change their market strategies during the experiment period cannot be measured due to the impact of competitors' strategies. Also, other factors like – product quality and after sales service may also influence the sales instead of advertising.

10.7.4 Media Research

Media selection is finding the most cost effective media to deliver the desired number and type of exposures to the target audience. The media planner has to know the capacity of the major media types to deliver reach, frequency and impact. The major media types are T.V., newspaper, radio, magazines or journals etc. Which media is to be selected – print or broadcast media this question is there to answer in the process of media selection.

10.7.4.1 Print Media

For many years advertisers have not been satisfied with circulation data as the sole yardstick for buying advertising media. It has been clearly demonstrated that the number of readers of a given copy of a publication not only significantly exceeds circulation but also varies from gross circulation figures as between various publications. An increasing amount of research has been applied in resent years to the measurement

of actual reading audiences. Techniques fro measuring reading audiences are now sufficiently perfected so that these data are generally accepted and widely employed. Newspapers and magazines are the most common types of print media. As the media circulation increases so, does the attractiveness of a newspaper or magazine to an advertiser, and the medium may raise the advertising rates. The best way to measure a publications' is through measuring its readership or total audience. Media buyers need to know the accurate circulation and readership figures to compare costs among various publications that reach similar audiences. For magazines the formula is:

$$\text{Cost per page per thousand circulation} = \frac{\text{Page rate} * 1000}{\text{Circulation}}$$

Newspapers advertising rates can be compared as:

$$\text{Cost pre line per million circulation} = \frac{10{,}00{,}000 * \text{line rate}}{\text{Circulation}}$$

= million rate in rupees

10.7.4.2 Broadcast Media

The most extensive research in connection with broadcasting media is the measurement of listening and viewing audiences. The following methods are commonly employed.

a. The recall method: The recall method obtains its data through listener surveys in which the respondent is asked to report the television and radio programs he heard during a specified period of time previous to the interview.

b. The diary method: In the coincidental diary method a form listing broadcasting stations and time periods is left with the respondent, who makes notations of programs received. The diaries are usually designed so that they can be attached to the radio or television set for convenient recording. The chief

advantage of this method is that the respondent records his listening or viewing at the time, thus eliminating the burden on memory. It also has the advantage of making it possible to obtain classification data regarding the listener or viewer. The primary weakness of this method is inaccuracies of reporting because there is a constant tendency to neglect to make entries at the time the set is actually in use.

c. The coincidental method: In this method investigators call telephone subscribers according to a predetermined sampling pattern. It provides more accurate data than other survey methods, since it eliminates the necessity for memory or entries in a diary. Also it is based on a report of actual viewing or listening rather than mechanical evidence that the radio set was operating.

d. The audimeter method: The audimeter is attached to TV or radio sets at random in the homes of potential consumers. The device maintains on a magnetic tape a continuous record of the set in usage channel on and the station channel to which tuned. Its advantage is that with the aid of audimeter one can accurately measure the set in use from a valid sample. Its disadvantage is that it does not indicate who is watching the set or listening to a radio set. It only records whether the set is tuned on to a particular channel or station. It tells nothing about the audience presence and their behaviour.

Check your Progress

4) What is a audimeter?

10.8 BRANDING REASEARCH

Branding research is conducted for products, services or people. It 1st determines attributes and benefits and a product in a specific target market. The information is obtained through qualitative research. The perception of each of their attributes is then tested quantitatively. This kind of research helps the SWOT analysis of a brand. The benefits of branding are:-

1) Customers get additional value from branding attributes for which they are willing to pay a police.

2) The company's gains through cost saving that is obtained from loyal customers. They also gain from the revenue generated by the added value. Further overall branding helps the company make optimum use of the companies yet uncovered potential, threats and opportunities that would rub off on the brand image.

3) Uncover the difference between desire and actual perceived image.

4) Maximize acquisition of new customer and market penetration.

5) Minimize loss of current customers.

6) Safeguard and increase the income revenue

7) Improve marketing and sales planning.

10.9 PACKAGING RESEARCH

Package graphics and copy are critical market variables in any product category particularly for non-advertised or under advertised brand. In self service shopping environment there is a greater likelihood of the customers reaching to design change in the existing package. These are 4 packaging research services that may be used:-

1) Package Screen - It involves the screening of 10-20 alternatives package designs and are objectively analyzed and 1 is finally chosen.

2) Package Check - It is an internet bared system where the representative sample is placed in the website and the respondent sees only one design and is asked to respond to a series of questions. This study is based on 75 completed interviews.

NOTES

3) Package Tests - This is a comprehensive internet based testing system where the representative sample from the internet panel is chosen and qualified respondents are invited to evaluate the package design.

4) Customer / Adhoc Design Package Research - This involves communication that is interactive the respondents are asked to read and interpret the package design, graphics and copy; they are then analyzed in an interview.

10.10 MARKET SEGMENTATION RESEARCH

Segmentation is concerned with identifying different groups of purchasers in a market in order to target specific products and services for each group or segment. By tailoring the offering (communication, product, channel, price) to different groups you are able to more precisely meet the needs of more customers and consequently to gain a higher overall level of share or profit from a market. The process of segmentation starts with research and market analysis to identify key segments. However the findings of the research are just the start. To be successful, implementing a segmentation strategy involves aligning the organisation to deliver appropriately for each segment and there are real business issues to be considered because each segment requires investment if it is to be properly addressed.

Check your Progress

5) What are the four approaches to segmentation research?

Segmentation is a major element of marketing. At a research-level there are four major ways of segmenting a market according to the level of precision you require and the type of data and analysis available about your customers. However, in finding different market segments it is important to keep in mind that the business will have to use the segments and implement segment specific business decisions - such as tailored products, pricing or service to meet the needs of each

segment. Important questions are therefore how are you going to place customers into each group and how are you going to target and track each group. Do you leave it up to the customer to select themselves into a segment, or do you have specific segment sales managers?

10.10.1 Apriori segmentation

A-priori (pre-existing) segments are the most basic way of creating market segments. In A-priori segmentation, the market is split according to pre-existing demographic criteria such as age, sex or social economic status. More sophisticated versions include lifestage (which combines information about age, presence of children and working status) and geodemographics.

A priori segments are easy to define and easy to target with advertising and media. For some sectors, for instance technology, there are such strong relationships between age and use, that a priori segments are all that are needed. However in other markets - for instance drinks, it is more difficult to use pre-existing variables for segmentation. A priori segmentations are also the simplest segmentation to apply and use. A database can be flagged or sorted on the pre-existing data and that data used to drive sales and marketing campaigns. However, although better than pure mass marketing, even the most sophisticated a priori systems are quite crude. In geodemographics there is the assumption that you buy or think the same way as your neighbour which is clearly not always the case.

10.10.2 Usage segmentation (also known as decile analysis or pareto analysis)

There are two ways of carrying out a usage segmentation, firstly customers are split according to their weight of use. –

heavy users/buyers being more important targets than light users. This segmentation can be carried out directly on customer databases and can be extremely powerful in focusing activity based on the value to the business, not just the number of contacts. Decile analysis splits users into 10 evenly numbered groups, which Pareto analysis splits the top 20% from the bottom 80%. This is normally used in business-to-business markets and is a core part of database analysis for consumer markets. Secondly, usage can be considered in terms of time and place. A cafe might sell sandwiches at lunchtime but main meals in the evening because the purchasers are looking for different things. It may even be the same purchaser just in a different "mode". Usage studies are also extremely common in market research, but normally to determine measures of market share and other metrics. However this information can also be used as the basis of a segmentation approach. Often usage segmentation is used to try and establish underlying driving forces from other demographic variables. So if women are more likely to be heavy users would it be easier to convert more light users who are female, rather than target their male counterparts. This focusing of market activity on groups that are similar to heavy users gives rise to measures such as "uplift" - the improvement possible over a purely random approach.

10.10.3 Attitudinal research and cluster analysis

When market research is used for usage studies, it is also often accompanied by attitudinal research - what do customers think or believe about the category in question. This is commonly achieved through banks of agree-disagree scales or ratings out of 5, 7 or 10. The aim of these studies is not just to understand commonalities in opinion, but also what makes one

group of users different from another. To understand how attitudes affect purchase statistical techniques such as "cluster analysis" are used where people with similar attitudes are combined together. For instance grouping those for whom the environment is important separately from those who think price is more important. This information can then be used to target groups by what they think and how they feel, rather than just who they are. This is particularly valuable in determining branding strategies and keeping a brand in tune with consumers. However, attitudinal clusters do not fit easily into database or conventional media targeting which are more often than not based on demographics. The translation from attitudes to demographics means that some of the usefulness of an attitudinal segmentation is lost. Companies can reach different attitudinal groups by offering a range of products and a range of communication, but clearly the lack of a clear definition means cross-over between the targeting of segments.

Attitudinal grouping also suffer from some problems with regard to their robustness and replicability. Cluster analysis cannot be carried out in the field so scoring systems (similar to credit scoring) or surrogate measures and variables are needed to allocate individuals to a group. These additional measures can be guessed at, but normally need to be defined and tested post-hoc. Repeating attitudinal analysis successful can be very difficult and expensive. Attitudinal groups may also change or move over time as some views become fashionable or unfashionable. It is possible to find a segmentation that quickly disappears or is superceded by events (imagine the music market). There is also debate about how attitudes change - is it the advertising and the product that create the attitudes, or do the attitudes lead to the choice of a particular product. In

particular a single individual in different circumstances or mode, may fit into a different segment. Capturing this complexity in a single dimensional study is difficult.

10.10.4 Needs based segmentation

The fourth method is to try and determine fundamental drivers for the decision to create what is known as a needs based segmentation. Most needs-based segmentation uses Conjoint Analysis to split a category into different levels of functional performance (see Conjoint Design). By understanding what elements are key drivers for individuals, specific needs and requirements can be identified from the trade-offs that each person makes. Using cluster analysis, this information can be drawn together to find different segments with similar preferences and needs from the product category in question. Needs based segments are typically the most actionable forms of segments as you know what drivers and performance the product or service has to satisfy. These are normally more stable than attitudinal groups as they should directly reflect and predict existing market share. However, like attitudinal studies, because cluster analysis is used, targeting each of the underlying groups can be difficult. Nonetheless you also have the benefit of being able to product a market model or market simulation using the Conjoint output.

10.10.5 Choosing a technique

The type of segmentation you use will depend on a lot of factors including the cost not only of conducting the research, but also of implementing the solution and the business impact. Consequently ideally for each segment or group you want to know what the economic value and the economic potential for each group is and have some idea as to whether this is increasing or falling. Consequently most quantitative segmentation studies

are detailed and complex. A more cost-effective approach is to develop groups based on qualitative research. Typically a business wants to minimise the number of segments it has as each costs money to target properly (database marketing and digital printing techniques allow for far finer targeting without too much additional cost). With small numbers of big segments, a good researcher will be able to identify these groups within a programme of qualitative research. This will not gather economic data, but it enables deeper insight into each group and, if monitored over time, provides core information about how segments change and develop.

10.11 SUMMARY

From the above discussions it would be clear that marketing research is heavily used in the marketing mix components such as Product Price, Place and Promotion

10.12 ANSWERS TO 'CHECK YOUR PROGRESS'

1. Recent applications of marketing research: Currently in the contemporary period marketing research is applied to competitive advantage, brand equity, customer satisfaction, and total quality management researches

a. Competitive Advantage: There has been a shift of focus in marketing, from delivering goods and services to consumers (satisfying their needs) to achieving a competitive advantage. Companies are embracing new tools, techniques, and strategies in order to remain competitive. This has resulted in a new agenda for marketing research in the nineties. Assessing competitive advantage can be done in a number of ways. The methods can be broadly classified as market-based assessment and process-based assessment. Market-based assessment is direct comparison with a few target competitors, whereas

process-based assessment is a comparison of the methods employed by the competitors in achieving their distinctive advantage.

b. Brand Equity: Brand equity is defined as a set of assets and liabilities linked to a brand that add to or subtract from the value of a productor service to a company and/or its customers. The assets or liabilities that underlie brand equity must be linked to the name and/or symbol of the brand. However, they can be usefully grouped into five categories:

1. Brand loyalty
2. Name awareness
3. Perceived quality
4. Brand associations in addition to perceived quality
5. Other proprietary brand assets: patents, trademarks, channel relationships, etc.

An appraisal of the brand based on the five dimensions involves addressing and obtaining answers to the questions that follow. Marketing research can help to provide answers to these questions. It is important to develop approaches that place a value on a brand, for several reasons.

First, since brands are bought and sold, a value must be assessed by both buyers and sellers. Which approach makes the most sense?

Second, investments to enhance brand equity need to be justified, as there always are competing uses for funds. A bottom-line justification is that the investment will enhance the value of the brand. Thus, some "feel" for how a brand should be valued may help managers address such decisions.

Third, the valuation question provides additional insight into the brand-equity concept.

At least four general approaches have been proposed to assess the value of brand equity.

One is based on the excess price that the name can command in the marketplace.

The second looks at how much it would cost to replace the brand with a new one.

The third is based on the stock price.

The fourth focuses on a brand's earning power.

c. Customer Satisfaction- In recent years business has become increasingly committed to the idea of customer satisfaction and product/service quality. The measurement of customer satisfaction and its link to product/service attributes is the vehicle for developing a market-driven quality approach. Customer satisfaction research has been around for a long time, but it has become a fixture at most large corporations only in recent years. The growth in the popularity of customer satisfaction research is, of course, a corollary to the quality movement in American business. The idea that the customer defines quality should not be new to marketers. However, its recognition in the Baldridge criteria has given this idea a credibility that was previously lacking. Satisfaction research, like advertising tracking research, should be conducted at planned intervals so as to track satisfaction over time. Thus, satisfaction research can be put in the context of an interrupted, time-series, quasi-experimental design. Over time, management will do various things to improve customer satisfaction, take measurements following these changes, and evaluate the results to see if the changes that were implemented had a positive effect on customer satisfaction. This approach requires a sequential research design that uses the results from each research phase to build and enhance the

value of subsequent efforts. During this process, it is imperative to study customers who were lost, to determine why they left. This issue must be addressed early in the system design. A useful step is to provide management with a framework for understanding, analyzing, and evaluating the status of customer satisfaction in the firm. A sequential design provides some level of comfort, because it allows for the luxury of making critical decisions after you have sufficient data to reduce the risk of error inherent in establishing a customer satisfaction system.

d. Total Quality Management- With foreign competition steadily eating away the profitability and the market shares of American companies, more and more of them are adopting total quality management (TQM) to become more competitive. TQM is a process of managing complex changes in the organization with the aim of improving quality. The first things on which a TQM company should decide are the guiding principles behind its data choices. Why these data, and not those data? As usual, the best rationale usually refers back to the bedrock of customer satisfaction. There should be a clear link between the kinds of data collected and maintained and the quality values of the company.

If short-term financial measurements drive the company, measures such as market value to book value and price-to-earnings multiples will dominate management reports and meetings. If, on the other, quality lies at the center of business strategy and planning, a larger share of the measurement and reporting will focus on quality issues. When companies are truly committed to quality values, many data issues resolve themselves. The power of measurements is clearly visible in applications of quality function deployment (QFD), a Japanese

import used to make product designs better reflect customer requirements. In QFD, a multifunctional team measures and analyzes in great detail both customer attitudes and product attributes.

Marketing research plays a crucial role at this stage of the process. Then the team creates a visual matrix in order to find ways to modify product attributes (engineering characteristics) so as to improve the product on the customer-based measures of product performance.

Along the way, the team must develop a series of measures of several different types. Marketing research is an invaluable part of QFD. Customer attributes are obtained through conjoint analysis or through other forms of survey research. Customer evaluations of competing products are also obtained through survey research. Hence, a thorough knowledge of marketing research is required.

2. Emerging applications of marketing research:

Some of the emerging applications of marketing research are in database marketing and relationship marketing.

a. ***Database Marketing-*** The power of database technology ties these three characteristics together. The database collects and analyzes customer information; it is able to target specific benefits to specific customers and it provides the means for accurately measuring results. With the advent of customer lists, direct marketing came into prominence. A database is a customer list to which has been added information about the characteristics and the transactions of these customers. Businesses use it to cultivate customers and develop statistical profiles of prospects most like their present customers—as they seek new customers. important, their response characteristics. With a database, marketers can use past actions by customers

to predict their future preferences or profile prospective customers for effective market segmentation. With a database, marketers can project additional sales—through cross-selling and repeat purchases. At the same time, by collecting information and learning more about current customers' tastes and preferences, marketers can effectively target new customers with the same characteristics and even predict the lifetime value of these newly acquired customers. In this way, the organization cannot just replace customers lost through attrition but can grow. A database should attempt to create:

- A unique identifier such as an ID or match code
- Name and title of individual and/or organization
- Mailing address, including ZIP Code
- Telephone number
- Source of order, inquiry, or referral
- Date and purchase details of first transaction
- Recency/frequency/monetary transaction history by date, dollar amounts (cumulative) of purchase, and products (lines) purchased

Credit history and rating (scoring)

Relevant demographic data for consumer buyers, such as age, gender, marital status, family data, education, income, occupation, length of residence at address given, geodemographic cluster information, and similar data of value

Relevant organization data for industrial buyers, such as standard industrial classification (SIC), size, revenues, number of employees, length of time in business, perhaps information about the area of the organization's economic or social location, and even information about the personality of individual buyers within the organization.

There are many ways to gather consumer data. You may use such direct means as surveys, questionnaires, and application forms, but you may also get information from secondary sources, such as credit-reporting bureaus and published directories. When you fill out an application for a rebate or a specially discounted or even free book or other item, you may wonder why some of the questions seem to be rather strange as far as the rebate or items are concerned. The firm is collecting personal data not normally available in any other medium. (Now you know!) If you apply for a free subscription to a trade publication you generally must respond to a rather lengthy set of questions, but that, too, provides data not usually available otherwise. You may use conventional advertising media to induce customers and prospective customers to call and write. One way to do this is to use a special form of inquiry advertising. And more than a few other companies are publishing free newsletters for customers today, encouraging letters from readers as an effective means of generating data for their databases. The last time you signed onto the Internet, someone could have watched what you did, what you said, or what you bought and then shared the information with a curious marketer who wanted to know. Powerful computers and high-tech scanners now enable marketers to monitor closely how, where, and when you spend your money. These electronic transactions speak much louder than words because they reflect actual behavior. It no longer matters what consumers say they do; marketers can now track what they really do. Most firms want to have a customer database and a prospect database. The customer database can categorize customers as active or inactive customers and inquiries.

Active customers: How recently have they purchased? How frequently have the purchased? How much did they spend? What are their product or service preferences? Identifying your most active customers can help you concentrate your resources on the most profitable segment of your customer list.

Inactive customers: How long have prior customers been inactive? How long had they been active? What was their buying pattern while active? What offers have they received since? This information can help you design promotions that re-activate your inactive customers.

Inquiries: From what media source did inquiries come? What was the nature and seriousness of the inquiry? Do you have any demographic or psychographic information on inquirers? Database enhancement can substantially increase the amount and quality of information you hold on each customer or prospect. In its simplest form, an enhancement might be the addition of age (from a driver's license record) or telephone number (from a directory record). Other possibilities include past transactions; demographic and psychographic data; credit experience, if pertinent; people on the move, evidenced by an address change; significant characteristics of a business; and a multiple of customer behavior and transaction data.

By overlaying multiple databases, you can eliminate duplication between and among the lists and identify "hotline names" (those who responded most recently) and "multi-buyers" (those who appear on more than one response list).Negative screening, such as a credit check, can be used to remove a record from a solicitation database.

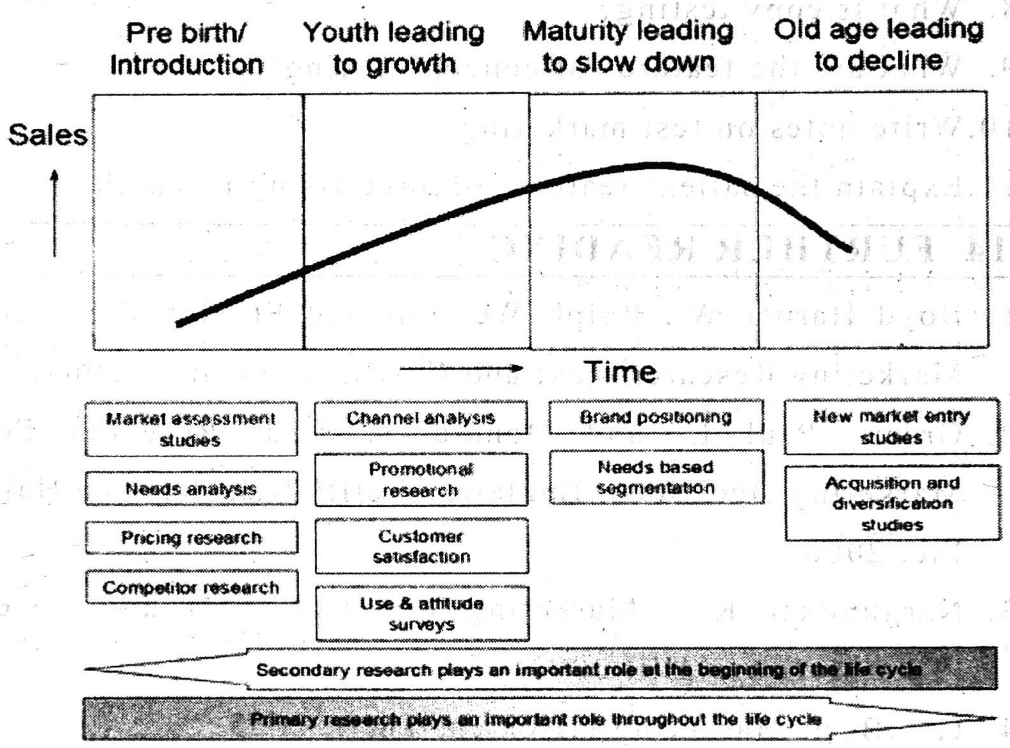
b. Relationship Marketing- The relationship marketing process incorporates three key elements:

1. Identifying and building a database of current and potential consumers, which records and cross-references a wide range of demographic, lifestyle, and purchase information

2. Delivering differential messages to these people through established and new media channels based on the consumers' characteristics and preferences

3. Tracking each relationship to monitor the cost of acquiring the consumer and the lifetime value of his or her purchases.

3. The Role of Market Research in a Product's Life Cycle



4. The audimeter is attached to TV or radio sets at random in the homes of potential consumers. The device maintains on a magnetic tape a continuous record of the set in usage channel on and the station channel to which tuned.

5. **Four approaches for segmentation research** are

1. Apriori segmentation
2. Usage segmentation
3. Attitudinal research and cluster analysis
4. Needs based segmentation

Applications of MR

NOTES

Applications of MR

NOTES

10.13 EXERCISES AND QUESTIONS

1. Explain the salient features of Product Research
2. What is meant by Price Research?
3. Explain the steps in Distribution Research
4. List out the various Promotion Research
5. Brief about Distribution research
6. What are the importances of Package Research?
7. List out the types of Segmentation research
8. What is copy testing?
9. What are the features of concept testing?
10. Write notes on test marketing
11. Explain the salient features of advertising research

10.14 FURTHER READING

1. Boyd Harper W., Ralph Westfall and Stanley F Stasch, Marketing Research Text and Cases, Irwin, Inc., 2007.
2. Green, Paul E, and Donald S. Tull, Research For Marketing Decisions, Englewood clifs N.J, Prentice Hall, Inc, 2006
3. Nargundkar R – Marketing Research Text and Cases (Tata McGraw- Hill 2002)
4. G.C.Beri, Marketing Research, TMH, 2008.